

.....
885 LysTyrAsnThrIleuThrValAsn...SerLeuSerGlyAsnGlySerH 900
2904 CCGCTTATGCGAACTCTTGGCTACCGCAAGCGGCAATGTAAGCTCG 2953
|
900 eTYTYLeuThrAspLeuSerAsnGlySerGlyAspGlyValValT 917
2954 CGGAAAGTTCGAAAGCACTTACACCTTGCTGCAACAATACCGCAAC 3003
917 hrLysSerAlaThrGlyAsnPhetHLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAGTCTCGAACAATGACGTAAGTGTGTAAGCAACACAC 3053
934 GluPro...AsnHisAsnGluLeuThrLeuPhaSpLa...SerLys 947
3054 ACCGCTGCTCGAAATCTTAATTCACCTGCAAAAGCAACACGTCGATG 3103
947 sAlaGlnArgAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGGCGCATGGCTTATCAGCTTATCGCAAAAGCGGCGGCTCGCTG 3153
964 euGlyAlaTrpLysTyrLysLeuArgAsnValAsnGlyArgTyrAspLeu 980
3154 CATATCCG..... 3162
981 TyrAsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnI 997
3162 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 3162
1014 snGluGluIleAlaArgValAspGluAlaProValProProAlaPro 1030
3162 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGluAsnSerLysGlnI 1047
3162 3162
1047 uSerLysThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaG 1064
3163GTCAAGAACAAGAG 3177
1064 InAsnArgGluValAlaLysGluAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTTCGCAAACTCGGCAAGCGGGA...GAACAGAGCGCGCTTGAC 3224
|
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
3225 GGCAAAAGCGCAACAATCGCGCAAAACAGCGGCAAAACAGCAACG 3274
|
1097 rThrGluThrLysGluThrAlaThrValGluLysGluLysAlaLysV 1114
3275 CGCAAGC.....CTTGACGCGCTGATGCG 3300
|
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130
3301 GCGGCGCAATGCGCAAGCAAGCAAGTGTGCGCAACGCGCGCG 3350
|
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147
3351 CGAGCGAGCGGGGAAATGCGGCAATGACGAGCGGAGAGAGAGAAA 3400
|
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlns 1160
3401 AACGGGTGAGCGGATTAAGACACCGCGCTTG...GCGAAACAGCGGAA 3447
|
1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
3448 GCG..... 3450

1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3451GAACCGCGCGGCTACCAACCGCGCTTC 3478
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrGln...P 1209
3479 CC.....CGCGCGCGCGCG 3492
|
1209 rOThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
3493 GCGCGCGGAGATTGCGG...CAACCGACCGCCCAACCGCAACCGCAAC 3539
|
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3540 GCAGGC.....GACCTGATCAGCGGTTATGCAATA 3571
|
1242 nAspArgSerThrValAlaIleuCysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTGATGATTTTCGCGC.....ACGCTCAACACG 3606
|
1259 LaValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCGCTACAGCAGATTTGACCGCGTG...TTGGCGGAAGACG 3653
|
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluG 1292
3654 CCGCAACGCGGCTTGACAGCGGCGATCGGGACCAACACTACCGT 3703
|
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSer 1309
3704 CGCAAGATTTCGCGCGCTACCGCCCAACAACCGACCTCGCCCAATCGGT 3753
|
1309 eSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAACCTCGGACGCGG...CGCGTCGATCGTTTCCCA 3800
|
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3801 CAACCGAGCGGAAACACTTCGACAGCGGCAATCGCAACTCGGACGCGC 3850
|
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThrIleuA 1359
3851 TTGCCAGCGTCCGTTTCGGGCAATCGGCATCGGCGAGCGTTCACATC 3900
|
1359 laGlnValAsnPhetYrSerLysTyrTyrAlaAspAsnIleTrpTyrLeu 1375
3901 GGCATCAGCGCGGCGCGGTTTGTAGTACGGCGACCTTCAGACGCGAT 3950
|
1376 GlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAsnH 1392
3951 CAGAGCAAAATCCGCGCGCGCGCTGCTCATTCAGCATTCAGCAAGAT 4000
|
1392 sAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaGlyL 1409
4001 ACCGCGAGGTTTCGCGCGATTCGCGCATCGAACCGCACATCGGCGACG 4050
|
1409 yAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyAlaArg 1425
4051 CGCTATTTCGTCCAAAAGCGATTCAGCATTCAGAAACGTCATATTCGC 4100
|
1426 TyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIleL 1442
4101 CACCGCGGCGCTTCGATTCACCGCTACCGCGGCGGCGCATTAAGCGGAT 4150
|
1442 sValAsnProIleSerValIleSerValThrAlaPheAlaGlnValAspLeuSerT 1459
4151 ATTCATTCAAACCGCGCAACACTTCATTCACGCGCTATTGACGCTG 4200
|
1459 yThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSerAla 1474
4201 TCCTATACCGATCCGCTTCGCGCAAGTCCGACCGCGTCAATACCGC 4250
|
1475 ArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGlyTy 1490


```
4251 CGATTGGCGCAGATTTCGGCAAAACCGCAGTCGCAATGGGGGCTAA 4300
      ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1490 rAspHeLaTyraSnaValGluAsnGlnGlnInIInTyraSnaIaGlyLeu 1507
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
4301 ACCCGCAATCAAGGTTTCACGCTGCTCCACGCTGCCCGCCCAAG 4350
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1507 ySeuLeuTyRHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
4351 GGGCCGCAATTGGAAGCGCAGCAGCGCGGCATCAATTGCGCTAC 4398
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1524 AlaLysGlnAlaGlnLysGlnLysThrAlaGlnLysLeuSerPhe 1539
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|

seq_name: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:PCT-US95-10661A-3

seq_documentation_block:
: Sequence 3, Application PC/TUS9510661A
: GENERAL INFORMATION:
: APPLICANT: Washington University, et al.
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10661A
: FILING DATE: 16-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,791
: FILING DATE: 25-AUG-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Treccarlin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: FP-59941/RFT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1541 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: PCT-US95-10661A-3

alignment_scores:
      Quality: 1614.00      Length: 1716
      Ratio: 1.732          Gaps: 48
Percent Similarity: 54.312 Percent Identity: 26.865

alignment_block:
US-09-303-518D-653 x PCT-US95-10661A-3 ..
Align seg 1/1 to: PCT-US95-10661A-3 from: 1 to: 1541

64 CGATTTCGCCCGGCTTACTAGCATATGCGTGTGCGGCAATCGCC 113
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
5 LysHeuLysLeuAsnPhelIleAlaLeuThrValAlaTyraIaLeuThr 21
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
114 CCAAGCCCGGGCGGACACTTATTTGGCATCACTACCAATACATC 163
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
21 oTyThrGluAlaIaLeuValArgAspAspValAspTyGlnIlePheA 38
164 CGAGCTTTGCCGAATAATAAGCAAGTTTGCGAGCGGGGCAAGATATT 213
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 rGAspPheAlaGlnLysGlnLysPheSerValGlyAlaThrAsnVal 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 GAGGTTTACACAAACAAAGGAGAGTTGTCGGCAATTCATGACGAAGC 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 LeuValLysAspLysAsnLysAspLeuGlyThrAlaLeuProAsnGly 71
264 C...CCGATGATTGATTTTCTGTGATG...CGTAAGCGGTGGCGG 307
   : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
71 yLleProMetIleAspPheSerValIaAspValAspLysArgIleAla 88
308 CATTCGGCGGCGATCAATATATGTCAGCGTGGCAAT...AACGGCGC 354
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
88 hrLeuIleAsnProGlnTyraValGlyAlaLysHisValSerAsnGly 104
355 TATAACAATGTTGATTTTGGT.....GGCGAGGGGAAGCA 389
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
390 TCCGATCACCACCGC.....TTTCTTACCAAAATTG 421
121 naLaLysAlaHisArgAspValSerSerGluGluAsnArgTyPheSer 138
422 TGAACAAATATATTAAGCAAGCACTAACCGC..... 456
138 aLgLuLysAsnGluTyProThrLysLeuAsnGlyLysThrValThrThr 154
457 .....CATCCTTATGGCGGCGATTCATATCATATCGGCTTGCA 494
155 GluAspGlnThrGlnLysArgArgGluAspTyTyMetProArgLeuAs 171
495 CAATTTGTGCAGATGCAGAACCTGTGAG...ATGACGATTATATG 541
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188
542 ATGGGTGGAATATACGCTGATTTAATTAATACCTGATGTGTGCAATC 591
188 sPaLaGlyThrTyraSnaSrgLnaSlnLysTyProLaPheValArgLeu 204
592 GGAGCAGCGCAGCAATATGGCGGCTGATGAAGACGA..... 630
205 GlySerGlySerGlnPheIleTyLysGlyAspAsnTySerLeuI 221
631 .CCCAATTAACCGCAAGTTCATATCAT.....ATTGCAAGCG 667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 eLeuAsnAsnHisGluValGlyLysAsnLysLeuValGlyAspA 238
668 CATATCTTGCGCTGCGTGGCGCAATACCTTTGSCACAAATGAGATAGGT 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
238 IaTyThrTyGlyIleAlaGlyThrProTyLysValAsnHisGluAsn 254
718 GGTGSCACAGCAACTTAGTAGCGCAAAATTAACATAGC...CCATA 764
   ::|::|::|::|::|::|::|::|::|::|::|::|::|
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspPro 271
765 TGSTTTTACCAACAGCA.....GSGTCATTTGGCGACA 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyAlaValIleGluAsp 288
800 GTGGCTCACCAGTGTATATGATGGCCCAAGCAAAAGCAAGTGAAT 849
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 eGlySerProLeuPheValTyraSrgGlnLysGlyLysTrpLeuPhe 304
850 AATGGGGATTCACAAACAGCAACCCATATAGCAAAAGCAATGCGTT 899
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
305 LeuGlySerTyAspPheTrpAlaGlyTyraSlnLysLysSer..... 318
900 CCAAGCTAGTTGCTAAAGATTG...TTCTATGATGAATCTTTGTGGAG 946
319 .....TrpGlnGluTrpAsnIleTyLysSerGlnPheThrLysA 332
```

```
947 ATACCATTCAGTATTCTACGAACACATCAATAATGGAATACTTTT 996
|||||
332 spValLeu..... 334
997 AAGACAAATATATATGCGAGGAAAAATGATCCAAATACACTA 1046
|||||
335 .....AsnLysAspSerAlaGlySerLeuIleGlySerLysThrSply 349
|||||
1047 TTCTTACCTTATAGATTAAACACGACGCTT.....CAAT 1084
|||||
349 rSerThrSerSerAsnGlySerThrSerThrIleThrGlyGlyLys 366
|||||
1085 TGTTTATGTTTCTTTCGAGACACACAGACAGACGCTTTATCATGCT 1134
|||||
366 erLeuAsnValAspLeuAlaAspGlyLysAspLysPro..... 378
|||||
1135 GCAGGTGGGGTCAACAGTTATTCGACCACGACTGAATATGAGAAATAT 1184
|||||
379 .....AsnHisGlyLysSerVa 384
1185 TTCTTATTCAGCAAGGAAAGTGAATTGATTAATACCAACACTCA 1234
|||||
384 lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnIleA 399
1235 ACCAAGGCGGCGGCTTTGATTGAGGATATTTACGATC...TCG 1281
|||||
399 spGlnGlyAlaGlyGlyLeuPheGluGlyAspTyrGlyValAlaGly 415
1282 CCTAAAAACAAGACGCTGGCAAGGCGGCGCTTCATATCACTGATG 1331
|||||
416 ThrSerAspAsnThrThrPlyGlyAlaGlyValSerValAlaGly 432
1332 CAGTACGCTTCTGGAAGTAACGCGCTGGCAAGACGCGCTGTC 1381
|||||
432 yLysThrValThrPlyValHisAsnProGlnTyrAspArgLeuAla 449
1382 AAATCGCAAGGACGCTGCTGCTCAAGCCAAAGGAGGAAAAACCAAGC 1431
|||||
449 yLysGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnGly 465
1432 TCGGTACGCTGGCGGCGGTAAGTCACTTATCAACGCGGAGCA 1481
|||||
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnG 482
1482 TCAAGCAAAAACAAGCCTTAGGAATCGCTGCTGACGCGCAGG 1531
|||||
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyAs 498
1532 GGACGCTCAACTGATGCGGATATATCACTCAACCCGACAACTCTAT 1581
|||||
498 erThrLeuValIleuAsnAspLysGlnValAspProAsnSerIleTyr 514
1582 TTTCGCTTTCGCGGCGGCTTTGATTGAAAGGCACTGCTTGGTT 1631
|||||
515 PheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACCGCATTCATAATACCATGAGAGGCGCATGTTGTCACCAACATC 1681
|||||
531 eAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548
1682 AAGACAAAGATCCACGCTTACCATTCACGCAATTAAGATATTACT... 1728
|||||
548 erThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564
1728 ..... 1728
565 ProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAsnPr 581
1728 ..... 1728
581 cTyrAlaPheArgArgIleLysAspGlyGlyLeuTyrLeuAsnLeuG 598

1728 ..... 1728
598 lAsnTyrThrTyrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614
1729 .....ACACCGGCATATAC..... 1743
615 LeuProLysAsnSerGlyGluSerAsnGluAsnThrLeuTyrMetGlyLys 631
1744 .....ACACCTTGATAGCAAA 1762
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnGluA 648
1763 AAGAAATTCCTTACACGCTTGGTTGGCGAAGATGACCAACAAACG 1812
|||||
648 rgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGluGly...LysAsn 663
1813 AACGGCGGCTCAATCTGATTAATACCAACCGGAGAGGATGACACTT 1862
|||||
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680
1863 ACTGCTTCGCGGAGAACAAATTTAAACGGCAATATCAACCAACAAACG 1912
|||||
680 erLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697
1913 GCAAACTGTTTTCAGCGGACAGACCGCACCGCCCTACATCATTTA 1962
|||||
697 lYThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGTGGTGCMAAATGGAAGT.....ATCCCAAGAGGAAT 2006
|||||
714 GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluA 730
2007 CGTGTGGCAACAGATTTGATGACCGCACATTTAAACGCAAACTTCC 2056
|||||
730 lValValGluAspAspTyrIleAsnArgAsnPheLysAlaIleThrIleMetA 747
2057 ATATTAGGCGGACAGACGCGGTGTTCC...CGCAATGTCGCAAGTG 2103
|||||
747 snValThrIleAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsnIle 763
2104 GAAGCGATTTGATTAAGCAATCAACGCCCAAGCAATTTGCTGCTCC 2153
|||||
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTyr.. 779
2154 ACCGCAATCAAGCCACACATCTGTACAGTTGCGGACTGGAGGCGTGA 2203
|||||
780 ....LysThrGlyAspThrValCysValArgSerAspTyrThrGlyTyrV 795
2204 CAGTTGTACCGAAACCAATTCACGACGATTAAGTCACTTGCCTTCATG 2253
|||||
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAAGACCGACATCAAGAGCAATGTACGCTTGGCATCGCTCATTT 2303
|||||
811 AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827
2304 AAATTCACAGACTTCCACACATCAACGCGCAATCTTACGACGCGGAG 2353
|||||
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTTACGCTTACGCGCAACGCCAACCCMAAGGCAACCTTCACG 2403
|||||
843 snSerGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGCAATGCCAAGCAACATTTAATCAAGCAATTAACGCA 2453
851 ..... 851
2454 CACATCGGCTTGGACATGCTTCATTATTAATCAAGCAACACGCGGTAC 2503
851 ..... 851
2504 AAAACGCACTGTGACGCTTTCGACACACGCTAAGCAACGTAAGCCAT 2553
```

```
851 ..... 851
2554 TCGGCACTCAAGCAATGTCTCCCTAGCGGATAGGCAATTCATTT 2603
851 ..... 851
2604 TGAACACAGCGCTTACCGGAAAAATCAGCGCGCAAGATACGGCAT 2653
851 ..... 851
2654 TACACTTAAAAAGACAGGAATGAGCGCTCCGTGGGCGACGAATTAGGC 2703
852 ..... SerHis1rPhis1leuThrGlyAsnSerAspValHis 863
2704 AATTAAACCTTGACAAAGCCACCATTACCTCAATTCCGCTATGACA 2753
864 GlnLeuAspLeuAlaAsnGlyHis1leHis1leuAsnSerAlaAspAsnSe 880
2754 CGATGGCGGAGCGCGCAACCGGCAAGTGGCGAGATGCCCGCGCGCC 2803
880 rAsn..... 881
2804 GTTCGCGCGCTCCCTATATTCGTTACGCCGCAACTTCGCGAGATCC 2853
882 ..... AsnValThr..... 884
2854 CGTTTCAACACGCTGACGGTAAACGGCAATTGACGGCTCAGGAACATT 2903
885 Lys1rAsnThrLeuThrValAsn...SerLeuSerClyAsnGlySerPh 900
2904 CCGGTTATGTCGGAACCTTCGCTACCGGCAAGGCAATTACAGCTGG 2953
900 eTyTyLeuThrAspLeuSerAsnLysGlnGlyAspLysValValValT 917
2954 CGGAAGCTTCGGAAGCAGCTTACCTTGGCTGTCAACATACCGGCAAC 3003
917 hTySerAlaThrGlyAsnPherThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAAGTCTCGAGCAATTGACGGTGTAGTGAGAGAAAACAAAC 3053
934 GluPro...AsnHisAsnGlnLeuThrLeuPheAspAla...SerTy 947
3054 ACCGCTGTCGGAATCTTAAATTACCTGCAAAACGAACAGCATGATG 3103
947 SALAGlnArAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGCGCATGCGCTTATCAGCTTATCCGCAAGACGGCGAGTCCGCTG 3153
964 euGlyAlaTrpLysTyTrLysLeuArGAsnValAsnGlyArGTyArAspLeu 980
3154 CATATATCG..... 3162
981 TyrAsnProGluValGluLysArGAsnGlnThrValAspThrThrAsnI 997
3162 ..... 3162
997 eThrThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 ..... 3162
1014 snGlnGluIleAlaArgValAspGluAlaProValProProAlaPro 1030
3162 ..... 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGlnAsnSerLysGlnI 1047
3162 ..... 3162
1047 uSerLysThrValGluLysAsnGlnGlnAspAlaThrGluThrThrLag 1064
3163 .....GTCAAGAACAAGAG 3177
||||| :::
```

```
1064 lAsnArGluValAlaLysGlnAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTCGCAACAACTCGGACAGCGGGA...GAACAGAGCGCGCTTGC 3224
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTCCCGCCCAACACAGGCGGAAAAAGACAAG 3274
1097 rThrGluThrLysGluThrAlaThrValGluLysGlnLysAlaLysV 1114
3275 CGCAAAAGC.....CTTGACGCCGCTGATTGCG 3300
1114 alGluThrGluLysThrGlnGluValProLysValThrSerGlnValSer 1130
3301 GCCGGGCGCAATGCCAGGAAAGCGAAGAGTGTGGCGGAACGGCGCG 3350
1131 ProLysGlnGluGlnSerGluThrValGlnProGlnAlaGluProAlaAr 1147
3351 GCAGCAGCGCGGGAATAATGCCGCTTATGACAGCGGAGGAGAAAGAAA 3400
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlns 1160
3401 AACGGTGCAGCGGAGTAAAGACACCGCCTTG...GCCAAACAGCGGAA 3447
1160 eGlnThrAsnThrThrAlaAspThrGlnGluProAlaLysGluThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3451 .....GAAACCGCGCGCTACACCGCCTTCC 3478
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCGCGCGC 3492
1209 rThrValAsnSerGluSerAsnLysProLysAsnArGHisArGArg 1225
3493 GCCCGCGCGGATTGCGG...CAACCGCAGCGCCCAACCGCAACCCCAAC 3539
1226 SerValArGSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3540 GCAGCGC.....GACGTGATCAGCGCTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuCyAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTGAGTGAATTTCCGCC.....ACGCTCAACAGC 3606
1259 lValLeuSerAspAlaArGAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCGTACAGGAGCAATTGGACGCGGTG...TTTGGCGAAGACCG 3653
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluI 1292
3654 CCGCAACGCGCTTGGACAAGCGGATCCGGGACACCAACACTACCGTT 3703
1292 yGlnTyAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrsers 1309
3704 CGCAGATTTCGCGCGCTTACCGCCAAACAGCAACGACCTGCGCAATCGGT 3753
1309 eSerGlnTyArGArGpHeSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAGAAAAACCTCGCGAGCGGG...CGCGTGGCGATCCTGTTTTCGA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3801 CAACGAGACCGGAAACACCTTGCAGACGCGCATCGCAACTCGGACGCGC 3850
1342 rValArGAsnSerAsnAsnPhAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCACAGGTCCGTTTTCGGGCAATACGCATCGGACGAGTTCCAGATC 3900
1359 lAcGlnValAsnPhenTySerLysTyTyTrAlaAspAsnHisTrpTyLeu 1375
```


612 CGCGCTGATGAGACGAGCAACCCATATACCGCAAGTTCATATCAT... 657
 221 pleuqlylsglglglgllyssersasp...Alaglyglytyrasnleul 237
 658ATTGCACCGCATATTCCTGGCTGCTGGTGGCAATACCTTTGCA 702
 237 ysleuvalglaysnalaatythrtytgcyllealeaglyhrrprotyrglu 253
 703 CAAAATGGATCAGGTGTGGCACAGTCACCTTAGTACC..... 741
 254 valasnhisglusnaspglyleuileaglypneglyasnserasnagl 270
 742GAAAAATTAACATFAGCCCATATGTTTTTTTAC 775
 270 utyrileasnprolysgluleuenserlysspro.....Leut 284
 776 CAACAGAGGCGTCATTTGGCGACAGTGGCTCACCAATGTTATCTAGAT 825
 284 hrasnryralaValleuagllyaspsercglyserProleuPheVallyrasp 300
 826 GCCCAAAAGCAAAAGTGGTTAATTAATGGGATTTGCCAAACGGCAACC 875
 301 ArgglulyselylystrpleuPheleuaglyserTyrraspyttrPalagl 317
 876 CTATATAGCAAAAAACAATGGCTTCCAGCTAGTGTAAAGATTTGTTCT 925
 317 ytyr.....AsnlyssersertpIngln 325
 926 ATGATGAATCTTGCTGGAGATACCATTACAGTATTCACAGAACACAT 975
 325 lutrpsnilletyrlysprogluPhealaglylsiletyrleuInlyr 341
 976 CAAATGGGAAATACTTTTTAAACGACATTAATTAATGGCGGAGAAAT 1025
 342SerallglySerle 346
 1026 CGATGGCAAAATAAACACATATCTCTACCTATAGATTAAAAACACAA 1075
 346 ulleglyserlystrAspyrSetrtipserseinsnglylstrhSert 363
 1076 CCGTT.....CAATGGTTAAATGTTTTTTATCCGACACACACA 1113
 363 hrlethrghrglyglulysserleuasnValaspLeuAlaspolylys 379
 1114 AGGAACTCGTTATCATCTGACAGTGGGGCTCAACACAGTTATCGACCAG 1163
 380 AspyrPro..... 382
 1164 ACTGAATTAATGGAGAAATATTCCTTTATTCACAAAGAAAGGTGAT 1213
 383AsnHISgllyssersValthrPhe.....Glueglyserglyhrrl 396
 1214 TGTACTTACCAACACATCAACCAAGGCGGCGGCTTTGTATTGTAG 1263
 396 euthrleuasnasnlnleasprngllylaaglyleuPhePheglu 412
 1264 GGTAAATTTTACGGTC....TCGGCTTAAAAACAAGAAAGTGGCAAGCGC 1310
 413 GlYAsPtyrGlVallylselylyhrrerkspsasnrrhrrtyrlysglyAl 429
 1311 GGGCGTTCAATTCAGTATGGCAGTACCGTTACTTGAAGTAACGGCG 1360
 429 aglyValserValaIagluglylystrValthrTyrrpyrsValInHnsAsp 446
 1361 TGGCAACGACGCGCGTCCAAATCCGCAAGGACGACGCTGCGTTCAA 1410
 446 roglIntyrrAsprrgleuAlalyslleaglylysglythrleuInleValglu 462
 1411 GCCAAAGGGAACCAACAGGCTCGGTCAACGGCGGACGCGTAAGTAT 1460
 463 GlYthrGlyAspAsnlyselyleuysValglYAspArglythrValIl 479
 1461 CTTAGATCAGCAGCGGAGCATCAAGGCAAAAAACAAGCCTTTATGTGAAA 1510

479 eLeuYsGInGInThrAsnGlySerGly...GlnHisAlaPheAlaSerGly 495

1511 TCGGCTTGGTCACGGCGACGGGGAGCGGTGCAACTGAATGCGGATATACG 1560

495 aGlyIleValSerGlyArgSerThrIleValIleuAsnAspAspGlyGln 511

1561 TTCACCCCGCAACAACACTGATATTTCGGCTTGGCGGTGGCGACGTTGGATT 1610

512 ValAspProAsnSerIleThrPheGlyPheArgGlyGlyAlaGlyLeuAspLe 528

1611 GAACGGGCAATTCGCTTTCCTTCACCGCATTCCTCAAAATACCGATGAAGGG 1660

528 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspGlnGlyAla 545

1661 CGATGATTTGCAACCCCAATTCACAGCAAAAGAATTCACCGTTACCATTTACA 1710

545 IaArgIleuValAsnHisSerThrSerLysHisSerThrValThrIleThr 561

1711 GCGAATAAGATATTATTCCTCAACCGCGCAAT 1740

562 GlyAspAsnLeuIleThrAspProAsnAsnValSerIleThrValVal 578

1740 1740

578 sProIleuGluAspAspAsnProThrAlaIleArgGlnIleLeuSerGlyTyr 595

1740 1740

595 YrGInLeuThrPheAsnGluGluAsnArgThrTyrTyrAlaIleuLysLys 611

1740 1740

612 AspAlaSerIleArgSerGluPheProGluAsnArgGlyGluSerAsnAsp 628

1741A 1741

628 nSerThrPleuTyrMetGlyThrGluLysAlaAspAlaGlnLysAsnAlam 645

1742 ACAACAACCTTGATAGCAAAAAGAATTCGCTTCACACGGTTGGTTGGC 1791

645 eTrsHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPheGly 661

1792 GAGAATGATGCAACCAACGACGCGGGCGGTGCAATCTGAATTACCAAC 1841

662 GluGluGluGly...LysAsnAsnGlyLysLeuAsnValThrPheLysGln 677

1842 GGAAGAAGCGGATCGCACATTCTACTGCTTCGCGGGCAACAATTTAAAG 1891

677 LysSerSerGluGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnG 694

1892 GCAATATTCACGCAACCAACGCGCAACACTGTTTTCAGCGCGACACGACA 1941

694 LysAspLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArgProThr 710

1942 CCGCAGCGCTTACAATCATTTAGAGAGCGGGGTCAAAAATGGAAGGT 1989

711 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspSerHis 727

1990 ...ATCCACAAGAGAAATCGTGTGGGACAAGACATGGATGCACGCCA 2035

727 sPheSerGluAsnAsnGluValAlaValAlaGluAspAspThrIleAsnArg 744

2036 CATTTAAACGCAAAACTTCATATTCAGGGCGCGACACGCGTGTTTCC 2085

744 snPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeuTyrSer 760

2086 ...CGCAATGTTCCCAAGATGGAAGCGGATGGCATTTAAACATCAACG 2132

761 GlyArgAsnValGluSerIleThrSerAsnIleThrAlaSerAsnAsnAl 777

2133 CCAAGCAGTTTTCGGTTGTCACCGCATCAAGCCACACCAATCTGTACAC 2182


```

3551 TGATCAGCCCTATGCAATAGCGGTTTGAGTGAATTTCCGCC.....3594
      |||
      |||
1256 eutHrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaValSala 1272
      |||
      |||
3595 .....ACGCTACAGCCGTTTCCGCTACAGCAGCAATTTGACCG 3635
      |||
      |||
1273 GluPheValAlaLeuAsnValGlyLysAlaValSerIlnHisIleSerGI 1289
      |||
      |||
3636 CGTG...TTTCCGAAGACCGCCGCAACGCCGTTTGACAGCGGATCC 3682
      |||
      |||
1289 nleuGluMetAsnAsnGluGlnTyrAsnValTrrPvalSerAsnThrS 1306
      |||
      |||
3683 GGGACACCAACACACTACCGTTCCGACAGATTTCCGCGCTACCGCAACA 3732
      |||
      |||
1306 erMetAsnLysAsnTyrSerSerSerIlnTyrArgPheSerSerLys 1322
      |||
      |||
3733 ACCGACCTGGCCAAATCGGTATGCAAGAAAACCTCGGACGCGGG...CG 3779
      |||
      |||
1323 SerThrGlnThrGlnLeuGlyTrrPaspGlnThrIleSerAsnValGI 1339
      |||
      |||
3780 CGTGCGCATCTGTTTTCGCAACACCGACCGCAACACCTTCGACGACG 3829
      |||
      |||
1339 nleuGlyGlyValPheThrTyrValArgAsnSerAsnAsnPheAspLysA 1356
      |||
      |||
3830 GCATCGCAACTCGGACGCGCTGCCACGCGTCCGCTTTGGGGCAATAC 3879
      |||
      |||
1356 latHrSerLysAsnThrLeuAlaGlnValAsnPheTyrSerLysTyrTyr 1372
      |||
      |||
3880 GGCATCGCGAGGTTCGACATCGCATCGACGCGCGCGGCTTTTACTAG 3929
      |||
      |||
1373 AlaAspAsnHisTrrPtyrLeuGlyIleAspLeuGlyTyrGlyLysPheGI 1389
      |||
      |||
3930 CGGACGCTTTCAGACGCGCATCGACGCAAAATCCGCGCGCGCTGTCG 3979
      |||
      |||
1389 nserLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406
      |||
      |||
3980 ATTAGCGCATTCAGGCAAGATACCGCGCAGTTTCGCGGATTCGCGCATC 4029
      |||
      |||
1406 lnpPheGlyLeuThrAlaGlyLysAlaPheAsnLeuGlyAsnPheGlyIle 1422
      |||
      |||
4030 GAACCGCATCGCGCGCAACGCGCTATTTGTCGCAAAACGCGATTAACG 4079
      |||
      |||
1423 ThrProIleValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439
      |||
      |||
4080 ATAGCAAAAGTCATATCGCATCGCACCGCGGCTTCGATTAACCGCTACC 4129
      |||
      |||
1439 aleuAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456
      |||
      |||
4130 GCGCGGCGCATTAAGCAGATTATTCATCAACCGCGCGCAACATTTCC 4179
      |||
      |||
1456 heAlaGlnValAspLeuSerTyrThrTrrHisLeuGlyIlu...PheSer 1471
      |||
      |||
4180 ATCAGCGCTTATTTGAGCGCTGTCTATACGATCGCGCTTCGCGCAAGT 4229
      |||
      |||
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnIlySerGI 1487
      |||
      |||
4230 CCGACGCGCGTCAATACCGCTATTCGCGCAGAGTTTCGCAAAACCC 4279
      |||
      |||
1487 yLysIleAsnValAsnGlyTyrAspPheAlaTyrAsnValGluAsnGlnG 1504
      |||
      |||
4280 GCAAGTCGGGAATGAGCGGTAAACCGCAAAATCAAGAGTTTCACGCTGCC 4329
      |||
      |||
1504 lnsGlnTyrAsnAlaGlyLeuLysLeuLysTrrHisAsnValLysLeuSer 1520
      |||
      |||
4330 CTCACGCTCGCGCGCGCAAGGGCGCGCAATTTGGAACCGCAGCAGCCG 4379
      |||
      |||
1521 leuAlleLysLeuTrrLysAlaLysGlnAlaGluLysGlnLysThrAl 1537
      |||
      |||
4380 GGGCATCAATTAAGCTAC 4398
      |||
      |||
1537 agLueLeuLysLeuSerPhe 1543
      |||
      |||

```

seq_name: /cgn2_6/ptodata/1/1aa/PTUS_COMP.pep:PCT-US95-10661A-4

```

seq documentation block:
: Sequence 4, Application PC/RTUS9510661A
:
: GENERAL INFORMATION:
: APPLICANT: Washington University, et al.
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10661A
: FILING DATE: 16-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,791
: FILING DATE: 25-AUG-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Treacartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: FP-59941/RPT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1545 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: PCT-US95-10661A-4

alignment_scores:
      Quality: 1587.00      Length: 1723
      Ratio: 1.710
      Percent Similarity: 53.860      Percent Identity: 26.291

alignment_block:
US-09-303-518D-653 x PCT-US95-10661A-4 ..

Align seg 1/1 to: PCT-US95-10661A-4 from: 1 to: 1545

64 CGCTTCGCGCGCTTACTTACCATATGCTGTGCTTCGCGCATTCCTGCC 113
      |||
      |||
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
      |||
      |||
114 CCAACCGCGCGGCGGACACACTTATTTGGGATCACTACCAATTAATCTAC 163
      |||
      |||
21 otyrThrGlnAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38
      |||
      |||
164 GCGACTTTCGCGAAATTAAGCAGATTTCGACGCGCGCGCAAGATATT 213
      |||
      |||
38 rAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
      |||
      |||
214 GAGGTTTACCAACAAAAGAGGAGTTGGTCGCGCAATGATGACGCAAGC 263
      |||
      |||
55 GluValArgAspLysAsnAsnArgProLeuGlyAsnValLeuProAsnGI 71
      |||
      |||
264 C...CCGATGATTGATTTTTCGTGTATCG...CGTACGCGCTGGCGG 307
      |||
      |||
71 ylleProMetIleAspPheSerValValAlaSpValAspLysArgIleAlaT 88
      |||
      |||

```



```
645 elasnHisIleasnAsnGluArgMetAsnGlyPheasnGlyTyrPheGly 661
1792 GAGAAAGATCGAACCAAAACGAGGGCGCTCAATCTGATTAACCAAC 1841
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
662 GluGluGluGly...LysAsnAsnGlyAsnLeuAsnValThrPheLysG1 677
1842 GGAAGAGCGGATCGCACTTACTGCTTCCGGCGGACAAATTTAAAG 1891
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
677 LysSerGluGlnAsnArgPheLeuThrGlyGlyThrAsnLeuAsnG 694
1892 GCATATACGCAACAAACGCGCAACTGTTTTCAGCGCGACAGCCGACA 1941
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 LysPheLeuAsnValGlnGlnGlyThrLeuPheLysSerGlyArgProTh 710
1942 CCCGACGCGCTACAAATCATTTAGGAGGGGTGTCAAAAATGGAAGT.. 1989
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
711 ProHisIleArgAspIleAlaGlyIleSerSerThrLysLysAspSerH1 727
1990 ....ATCCCAAGAGAGAAATCGTGTGGACAAACGATTTGGATCGACGCA 2035
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
727 sPheSerGluAsnAsnGluValValValGluAspAspTPR1LeuAsnArgA 744
2036 CATTTAAAGCGGAAACTTCATATTCAGCGCGGACAGACGGGTGTTCC 2085
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
744 snPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeuTyrSer 760
2086 ...CGCAATGTGCAAAAGTGAAGGAGATTGCAATTTAAGCAATCAGC 2132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
761 GluArgAsnValGluSerIleThrSerAsnIleThrAlaSerAsnSna1 777
2133 CCAAGCAGTTTGGGTGTGCGACCGCATCAACACACAACTGTACAC 2182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
777 aluYsValHisIleGlyTyr....LysAlaGlyAspThrValCysValA 792
2183 GTTCGACGTGACGGCTGTACAGATTGTACGAAAAAACCATTAACCGAC 2232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 rgsrAspTyrThrGlyTyrValThrCysThrThrAspLysLysSer... 807
2233 GATAAAGTGTGCTTCATGTAGCAAGACCGACATCAGAGGCAATGTAC 2282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
808 AspLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnValAs 824
2283 CCTTGGCGATCAGCGTCATTTAAATCTCACAGGACTTGCACACTCAAG 2332
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
824 nLeuThrGluSerAlaAsnPheValLeu..GlyLysAlaAsnLeuPheG 840
2333 GCATCTTAATGCAAGCGGAGACAGCAGCTATACGTTACGGCGCAACGC 2382
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 LysThrIleGlnSerArgGlyAsnSerGlnValArgLeuThrGluAsn... 855
2383 ACCCAAAAGCGCAACCTCAGCCTGTGGGCAATGCCCAAGCAACATTAA 2432
855 ..... 855
2433 TCAAGGCACATTAAACGCAACACATCGGCTTCGACAAATGTTCAATTA 2482
855 ..... 855
2483 ATCTAAGCAACAGCCGCTACAAACGCGAGTGTACGCTTTCGACAAAC 2532
855 ..... 855
2533 GCTAAGCAAAAGCTAAGCCATTCGCACTCAACGGAATGTCTCCTTACG 2582
855 ..... 855
2583 CGATAAGCAGATTCCATTTTGAAACAGCCGCTTACCGGAAAAATCA 2632
855 ..... 855
2633 GCGGCGCAAGGATACGGCATTTACATTAAAGACAGCGAATGACGGCTG 2682
856 .....SerHisTrpHisLeu 860
```

```
2683 CCGTCGGCGACGAATTAAGCAATTTAAACCTTGACAAACGCAACCATTAAC 2732
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleH1 877
2733 ACTCAATTCGCCCTATCGACAGATGCGGACGCGCAACCGGCACTG 2782
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
877 sLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CCGCAGATGCGCGCGCGCGCTTCGCCGCTTCCCTATTATCGTTACG 2832
886 .....AspValThr 888
2833 CCGCCACACTTCGCGAGAAATCCGCTTCAACACGCTGACGTTAAGCGCA 2882
889 .....LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAACGGTCAAGGACATTCGCTTATGTGCAACCTTCGCTAC 2932
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
897 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 914
2933 GCAGCGGCAATTTGAAGCTGGCGGAAAGTTCGAAAGCAGCTTACACCTTG 2982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
914 lnglyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 930
2983 GCTGTCAACAATACCGGACAGCAACCGCTAAGTCTCGCAATGAGCGT 3032
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
931 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLe 946
3033 AGTGAAGAGAAAGACAAACACACCGCTGTCCGAAATCTTATTTACACC 3082
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
946 uPheAspAla....SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TCCAAAGACACACGTCGATGCGCGCATGCGCTTACCTTACCTCCG 3132
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
961 euValGlyAsnThrValAspLeuGlyAlaTyrLysLysLeuArgAsn 977
3133 AAAGCGGAGGATCCGCTGCATTAATCCG..... 3162
978 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnG1 994
3162 ..... 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 ..... 3162
1011 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027
3162 ..... 3162
1028 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1044
3162 ..... 3162
1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061
3162 ..... 3162
1061 sPAlaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1077
3163 .....GTCAAAGAACAGAGAGCTTCGACAAACCTCGGCAAGGGGGA.. 3204
1078 SerAsnValLysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlySe 1094
3205 GAACAGAGGCGCGCTTGACGCGCAAAACAGCACAACTTCGCGCAAC 3253
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1094 rclurThrLysGluThrGlnThrThrGluThrLysGluThrAlaThrValG 1111
3254 AACAGCGGAAAAAGACAACGCGCAAGC..... 3282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1111 luluYsGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127
```

```

3283 ...CTTGACGGCGTGTATGCGCGCGGCAATGCAACGCAAAAGGAGA 3329
      ::::: ::::: ::::: :::::
1128 LysValAlThrSerGlnValSerProLysGlnGlnSerGlnThrValGln 1144
      ::::: ::::: ::::: :::::
3330 AAGTTTCCCAACCGCGCGCGGCAAGCGGGGAAATGCCGGCATTA 3379
      ::::: ::::: ::::: :::::
1144 nProGlnAlaGlnProAlaArg.....GluAsnAspProThrV 1157
      ::::: ::::: ::::: :::::
3380 TGCAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3429
      ::::: ::::: ::::: :::::
1157 AlaSerGlnLeuGlnProGlnSerGlnThrValAlaSerPheGln 1173
      ::::: ::::: ::::: :::::
3430 TTG...GCGAAACAGCGCGGAGCG..... 3450
      ::::: ::::: ::::: :::::
1174 GlnProAlaLysGlnThrSerSerAsnValGlnGlnProValThrGln 1190
      ::::: ::::: ::::: :::::
3451 .....GAAACCC 3457
      ::::: ::::: ::::: :::::
1190 rThrThrValAsnThrGlnAsnSerValValGluAsnProGlnAsnThr 1207
      ::::: ::::: ::::: :::::
3458 GCGCGGCTACACCGCGCTTCCCG..... 3480
      ::::: ::::: ::::: :::::
1207 hrProAlaThrThrGln...ProThrValAsnSerGlnSerSerAsnLys 1222
      ::::: ::::: ::::: :::::
3481 .....CGCGCGCGCGCGCGCGCGGATTTGCGG...CAACGCA 3518
      ::::: ::::: ::::: :::::
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValGln 1239
      ::::: ::::: ::::: :::::
3519 GCGGCAACCGGCAACCGGCAACCGG.....GACC 3550
      ::::: ::::: ::::: :::::
1239 nProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuGlnAspL 1256
      ::::: ::::: ::::: :::::
3551 TGATCAGCGCTTATGCCAATAGCGGTTGAGTGAATTTCCCGC..... 3594
      ::::: ::::: ::::: :::::
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAla 1272
      ::::: ::::: ::::: :::::
3595 .....ACGCTCACAACGCGTTTTCGCGCTACAGAGCAAGTTGACCG 3635
      ::::: ::::: ::::: :::::
1273 GlnPheValAlaLeuAsnValGlnLysAlaValSerGlnHisIleSerGln 1289
      ::::: ::::: ::::: :::::
3636 GGTG...TTTGCGGAGAGCGCGCGCAACGCGGTTTGACACAGCGGCATCC 3682
      ::::: ::::: ::::: :::::
1289 nLeuGlnLeuAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1306
      ::::: ::::: ::::: :::::
3683 GGGACACCAACAACTACCGTTCGCAAGATTCGCGGCTACCGGCAACAA 3732
      ::::: ::::: ::::: :::::
1306 erMeAsnLysAsnLysSerSerSerGlnGlnArgAlaPheSerSerLys 1322
      ::::: ::::: ::::: :::::
3733 ACCGACCTGGCGCAATCGGTATGCAAGAAACCTGGCGAGCGG...CG 3779
      ::::: ::::: ::::: :::::
1323 SerThrGlnThrGlnLeuGlnLysPheGlnThrIleSerAsnAsnValGln 1339
      ::::: ::::: ::::: :::::
3780 GGTGCGGCACTCGTTTTCGCAACCGGAGCGGAAACCTTCGACAGCG 3829
      ::::: ::::: ::::: :::::
1339 nLeuGlnLysAlaPheThrValArgAsnSerAsnAsnAsnAsnLysAla 1356
      ::::: ::::: ::::: :::::
3830 GCATCGGCAACTCGGCGGCTTCGCCACGGTTCGCGTTCGCGCAATAC 3879
      ::::: ::::: ::::: :::::
1356 IaThrSerLysAsnThrLeuAlaGlnValAsnPheLysSerLysTyrTyr 1372
      ::::: ::::: ::::: :::::
3880 GGCATCGGCGAGGTTCGACATCGGATCGCGGCGCGGCGGTTTGTAGTAG 3929
      ::::: ::::: ::::: :::::
1373 AlaAspAsnHisTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1389
      ::::: ::::: ::::: :::::
3930 CGGCAAGCTTTCAGACGCGCATCAGAGGCAAAATCCCGCGCGCTGCTGC 3979
      ::::: ::::: ::::: :::::
1389 nSerLysLeuGlnThrAsnHisAsnAlaLysPheAlaArgHisThrAlaG 1406
      ::::: ::::: ::::: :::::
3980 ATTACGATTCAGGAGATACCGCGGCGGATTCGCGGCGGATTCGCGGATC 4029
      ::::: ::::: ::::: :::::
1406 nPheGlnLysThrAlaGlnLysAlaPheAsnLeuGlnLysPheGlnLys 1422
      ::::: ::::: ::::: :::::
4030 GAACGCGCATCGGCGCAACGCGGTATTCCTCCAAAGGCGGATTCACG 4079

```

```

1423 ThrProIleValGlnLysAlaArgTyrSerTyrLeuSerAsnAlaAspPheAl 1439
      ::::: ::::: ::::: :::::
4080 ATACGAAACGTCATATATGCGCAACCGCGGCGCTTGATTCACACCGGTACC 4129
      ::::: ::::: ::::: :::::
1439 alaAspGlnAlaArgIleLysValAsnProIleSerValLysThrAlaP 1456
      ::::: ::::: ::::: :::::
4130 GCGCGGCGCATTAAGCAGATTAATTCATCAAAACCGCGCAACATTTCC 4179
      ::::: ::::: ::::: :::::
1456 heaLacInValAspLeuSerTyrThrHisLeuGlnGln...PheSer 1471
      ::::: ::::: ::::: :::::
4180 ATCAGCGCTTATTAAGCCTGTCTTATACCGATGCCGCTTCGCGCAAGT 4229
      ::::: ::::: ::::: :::::
1472 ValThrProIleLeuSerAlaArgTyr...AspAlaAsnGlnGlnSerGln 1487
      ::::: ::::: ::::: :::::
4230 CGGACGCGCGGTCAATACCGCGGTATTCGCGAGATTCGCGCAAAACCC 4279
      ::::: ::::: ::::: :::::
1487 LysLysIleAsnValAsnGlnTyrAspPheAlaTyrAsnValGluAsnGln 1504
      ::::: ::::: ::::: :::::
4280 GCAATGCGGATAGGCGGTAAACGCGCAATCAAGTTTCAGCGGTGCC 4329
      ::::: ::::: ::::: :::::
1504 nGlnThrAsnAlaGlnLysLeuLysLeuLysTyrHisAsnValLysLeuSer 1520
      ::::: ::::: ::::: :::::
4330 CTCACGCTGCGCGCGCAAGGCGCGCAATTCGAACGCGACACACGCGC 4379
      ::::: ::::: ::::: :::::
1521 LeuIleGlnGlnLysThrLysAlaLysGlnAlaGlnLysGlnLysThrAl 1537
      ::::: ::::: ::::: :::::
4380 GGCATCAATATGAGCTAC 4398
      ::::: ::::: ::::: :::::
1537 aGlnLeuLysLeuSerPhe 1543

```

seq_name: /cgn2_6/ptodata/1/lae/6B_COMB.pep:US-08-296-791-5

seq_documentation_block:

Sequence 5, Application US/08296791

Patent No. 6245337

GENERAL INFORMATION:

APPLICANT: St. Gene III, Joseph W.

TITLE OF INVENTION: Haemophilus Adherence and Penetration

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296-791

FILING DATE: 25-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Treacartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1702 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-296-791-5

alignment_scores:

Quality: 1581.50 Length: 1867
 Ratio: 1.677 Gaps: 48
 Percent Similarity: 50.509 Percent Identity: 25.174

alignment_block:

US-09-303-518D-653 x US-08-296-791-5 ..

Align seg 1/1 to: US-08-296-791-5 from: 1 to: 1702

```

64 CGCTTCGCCGCCCTTACTTATGACCATATGCTGTGCTGGCATTCGCC 113
   :::::|||||: :::::|||||: :::::|||||: :::::|||||:
5  LysPheLeuSLeuAsnPhelLeaLeuThrValAlaIalValLeuThrPr 21
114 CCAAGCCCGCGGCGACACACTATTTCGGCATCACTACCATCTACTATC 163
   | :::::|||||: :::::|||||: :::::|||||: :::::|||||:
21 oYrThnGlAlaIalLeuValArGAspAspValAspTyGlnIlePhea 38
164 GCGACTTGGCGAAATAAAGCAAGTTGCACTGCGGCGAAAGATATT 213
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 rGAspPheAlaGlAsnIlySgIlyArGpHeSerValIglYAlaThrAsnVal 54
214 GAGGTTTACACAAAAGGGAGTTGTCGCGCAAAATCGATGACGAAGC 263
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
55 GluValArGAspIlySAsnAsnHisSerLeuGlyAsnValLeuProAsnG 71
264 C...CGGATGATGATTTTCTGTGTATCG...CGTAACGGCGTGGCGG 307
   | :::::|||||:|||||:|||||:|||||: :::::|||||:
71 YIleProMetIleAspPheSerValAlaAspValAspIlyArgIleAlaT 88
308 CATTTGGCGGCGCATATATATGAGCGTGCGACAT...AAGCGCGC 354
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
88 hrLeuIleAsnProGlnIlyrValValIglYAlaIlyshIstValSerAsnGly 104
355 TATAACAATGTTGATTTTGGT.....GCGGAGGGAAGCAA 389
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
390 TCCCGATCGACCGC.....TTTCTTACCAATG 421
   | :::::|||||: :::::|||||: :::::|||||: :::::|||||:
121 nAspIlySerHisArGAspValSerSerGluGluAsnArGlyrPheSerV 138
422 TGAAGAAATAATATAAAGCAGGACTAACGCG..... 456
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 AlGluIlySAsnGluIlyrProThrIlyLeuAsnGlyIlySalValAlaThrThr 154
457 .....CATCCTTATGGCGCGATATATCATATGCCGCGTTGCA 494
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 GluAspGlnThrGlnIlySArGArGAspTyGlyrTyMetProArGLeuAs 171
495 CAATTTGTCACAGATGCAAGACCTGTGAG...ATGACCATTAATATG 541
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 pIlySPrheValThrGluValAlaProIleGluAlaSerThrAlaSerSerA 188
542 ATGGGTGGAATACGCTGATTAATAATACCTGATGCTGTTCGAATC 591
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 sPlalagIlyrThrIlyAsnAspGlnAsnIlystyrProAlaPheValArGLeu 204
592 GGAAGCAGGCAACAATATGGCGGCTGTGATGAAGCAGAA..... 630
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 GlySerGlyThrGlnIleThrIleTyIlySgIlyAspAsnTySerLeuI 221
631 .CCCAATTAACCGGAAAGTTCAATATCAT.....ATTGCAAGG 667
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
221 eLeuAsnAsnHisGluValAlaGlyIlyAsnAsnLeuIlySLeuValGlyAspA 238
668 CATATTTCTGGCTGTGCTGGCAATACCTTTCGCAAAATGATGATCAGT 717
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 lAtyrThrTyGlyIleAlaGlyThrProTyIlyValAsnHisGluAsn 254
718 GGTGGCAAGTCAACTAGTAGCGAAAAAATTAAACATAGC...CCATA 764
   :::::|||||: :::::|||||: :::::|||||: :::::|||||:

```

```

255 AsnGlyLeuIleGlyPheGlyAsnSerIlySgIlyHisSerAspProIy 271
765 TGGTTTTTACCACAGGA.....GGCATTTGGCGCA 799
   |||||:|||||: :::::|||||: :::::|||||:
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyAlaValLeuGlyAspS 288
800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAACTGTTAAT 849
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
288 eGlySerProLeuPheValIlyrAspArGluIlySgIlyIlyStrPhePhe 304
850 AATGGGATATTCGCAACAGCAACCCCTATATAGAAAACCAATGGCTT 899
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 LeuGlySerTyArSpPheThrAlaGlyTyAsnIlySlySerTrpGlnI 321
900 CCAGCTAGTCTTAAGATGTTCTATGATGAATCTTCTGTGAGATA 949
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
321 uTPAsnIleTyIlySProGluPheAlaIlyThrValLeuAspIlySAsp 338
950 CCCATTCAGTATTCACGACCCATCAAAATGGGAATACTTTTATAC 999
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 hr..... 338
1000 GACAAATAATATGCGCGAGAAAATCGATGCCAAACATAACTATTC 1049
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 .....AlaGlySerLeuIleGlySerAsnThrGlnTyAs 350
1050 TCTACTTATAGATTAATAACAGCAACCGTT.....CAATGT 1087
   | :::::|||||: :::::|||||: :::::|||||: :::::|||||:
350 nTPAsnProThrGlyIlySThrSerValIleSerAsnGlySerGluSerL 367
1088 TTAATGTTCTTATCCGACAGCAGACAGAACCTGTTATCATGCTGCA 1137
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGTCACAGTTATCGACCCAGACTGAATAAGAGAAATATATTC 1187
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 .....ThrAspSerIlySAsnAsnHisGlyIlySerValTh 390
1188 CTTTATGCAAAAGAAAGATGATTTACTTACCAAGCAACTCAACC 1237
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
390 rLeu.....ArgIlySerGlyThrLeuThrLeuAsnAsnIleAspG 405
1238 AAGCGCGGCGGTTTGTATTTTGGAGGTAATTTTACGCTC...TCGCT 1284
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 lngIlyAlaGlyIlyLeuPhePheGluGlyAspTyrGluValIlyGlyIly 421
1285 AAAAACAAGAACGTGGCAAGCGCGGCGCTTCATATCAGTATGTCAG 1334
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 SerAspSerThrThrTrpIlyGlyAlaGlyAlaSerValAlaAspGly 438
1335 TACGCTTACTTGGAAGTAACCGCGCTGCAAAAGACCGCTGCCAAA 1384
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 sThrValThrTrpIlySValHisAsnProIlySerAspArgLeuAlaIys 455
1385 TCGGCAAGGCAAGCGCTGCTGTTCAAGGCAAGGGAAGAAACAGGCTCG 1434
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 lGlyIlyGlyIlyThrLeuIleValIglIlySgIlyGluAsnIlyGlySer 471
1435 GTCAGCGTGGCGAGCGGTAAGTCATCTTAGATACGACGCGGACATCA 1484
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
472 LeuIlySValIglYAspGlyThrValIleLeuIlySglnGlnAlaAspAlaAs 488
1485 AGGCAAAAACAAGCCTTAGGAATAGGCTGTGACGCGCGGAGGGA 1534
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
488 nAsnIlySValIlySalAspSerGlnValGlyIleValSerIlyArGSer 505
1535 CGGTGCAACTGAATGCCGATATCATGTTCAACCCCGCAAACTATATTC 1584
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
505 hrValIleuAsnAspAspIlySglnValAspProAsnSerIleTyPhe 521
1585 GGTTCGCGGCGGAGCGTTTGGATTTTGAACGGCGCATTCGCTTGTTCA 1634
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
522 GlyPheArGlyIlyArGLeuAspAlaAsnGlyAsnAsnLeuThrPheG 538

```

1635 CCGCATCAAAATACGATGAGGGGCGATGTTGCAACACATCAAG 1684
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
538 UHSHLEARGSNILEASPSGLYALARGLEVALASNHISNTHRS 555
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1685 ACAAGAATCCCGCTTACGATTCAGCAATAAAGATATTACT 1728
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
555 ERLYSTHRSERTHVALTHRLERHRLGYSERLEULETHASPRO 571
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1728 1728
572 ASNTHRLERHRLPROTYRASNILEASPALAPROASPSIUSPSPNPROLY 588
1728 1728
588 RALPHEARGARGILEYASPSGLYGLINLEUTYLEUASNLGUA 605
1728 1728
605 SNLYRTHRYTALALEUARGYSGLYALASERTHARGSERGLULEU 621
1729ACACCGCGCATTAAC..... 1743
622 PROLYSASNISERGLYUSERSNGLUASNTRIPLEUTYRMETGLYSTH 638
1744AACACTTGATAGCAAAAAG 1765
638 RSETPALALALALYSARGASNVALMETASNHSILEASNSNGLUARGM 655
1766 AAATGGCTACACGCTGGTGGCGAGAAAGATGCAACCAAAAAGAC 1815
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
655 ETASNGLYPHEASNGLYTRPHEGLYGLUGLUGLUGLY..LYSASN 670
1816 GGGGCGCTCATCTGAATTACCAACCGGAAGACGGATCGCACTTACT 1865
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
671 GLYSNLEUASNVALTHRPHELYSGLYLSERGLUGLINSNARGPHELE 687
1866 GCTTCCGCGGAGCAAAATTTAAACGCAATATCAACCAAAAAGACGCA 1915
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
687 ULEUTHRLGLYGLYTHRSNLEUASNGLYASPLEULYSVALGLULYSGLY 704
1916 AACTGTTTTCAGCGGACGACCGACCGACGCGCTACATCATTTTGA 1965
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
704 HRLPEUHLERSEGLYARGPROTHRPHONISALARGASPLLEAGLY 720
1966 AGCGGCTGCTAAATAATGGAAGT....ATCCACAGAGAGAAATCGT 2009
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
721 ILESERSETHRLYSASPGLNHSIPHEALAGLUASNSNGLUVALA 737
2010 GTGGGACACGATTCGATCGACGACATTTAAACGGAATACTTCATA 2059
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
737 IVALGLUASPSRTPRILEASNARGASNPHELYALATHRSNLEASN 754
2060 TTCAGGCGGACANAGCGGTGCTTC...CGCAATGTTGCCAAGTGAA 2106
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
754 ALTHRSNLSNALATHRLERYSERGLYARGASNVALAALSNIETHR 770
2107 GCGCATTCGCTTAAGCAATCAGCCCAACGACTTTTCGTCGCAAC 2156
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
771 SERASNILETHRALASERASPSNALSVALHSILEGLYTR.... 785
2157 GCATCAAAACACACATCTGTACACGTTGCGACGAGCGGTGCACA 2206
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
786 LYSALAGLYSPHRYVALCYSVALARSERASPTYRTHGLYTRYALT 802
2207 GTTGTACGAAAAAACCATTAACGAGATAAAGTATGCTTCATTGAGC 2256
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
802 HRCYSTHRLTHRASPLYSLEUSER...ASPLYSALALEUASNSERPHEASN 817
2257 AAGACCGACATCAGAGCAATGTCAGCTTCGATCAGCGTCAITTTAA 2306
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
818 ALATHRSNVALSERGLYASNVALASNLEUSER..... 828

2307 TCTCAAGAGACTTGCACACATCAACGGCAATCTAGTCAGGCGGAGACA 2356
828 828
2357 CGCACTATACGTTACGGCGCAAGCCACCACCAACGCACTCAGCCTC 2406
828 828
2407 GTGGCAATGCCCAAGCACATTTAATCAAGCCACATTTAAACGGCAAC 2456
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
829 ...GLYSNALSNSNHEVALLEUGLYLSALASNLEUPHEGLYTHRL 844
2457 ATCGGCTTCGACAAATGCTTCATTTAATCTAAGCAACAGCCGTACAA 2506
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
844 ESERGLYTHRLYASNSERLNVALARGLEUTHRLGUSN..... 857
2507 ACGGCACTGACGCTTCCGACAAACGCTAAGGCAACGTAAGCATTC 2556
857 857
2557 GCACCTCAAGGCAATGTCCTCCCTAGCCGATAGGAGATTCATTTGA 2606
857 857
2607 AAACAGCCGCTTACCGGAAAAATCAGCGCGGCAAGATACGGCATTA 2656
857 857
2657 ACTTAAACAGCGCAATGAGCGCTGCCGCGGACGCAATTAAGCAAT 2706
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
858SERHSTRPHLSLEUTHRLGLYSPSERASNVALASNGLN 870
2707 TTAACCTTGACACAGCCACCATTAACATCAATTCGCCCTATGACAGA 2756
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
871 LEUASNLEASPSYSELYHSILEHSLEUASNALA....GLIASN 885
2757 TGGGCGAGGCGGCAACCGGCAAGTCGCGAGATGCGCGCGCGCTT 2806
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
885 PALASNLYSVALTHRLTHR..... 891
2807 CGCGCGCTTCCTATTATTCGTTAGCGCGCACTTCGGCAGAAATCCGT 2856
891 891
2857 TTCACAGCGTGCAGGTAACGCAATTAAGCGTCAGAGCAATTCG 2906
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
892 TYRASNTHRLERHRYVALASN...SERLEUSERGLYASNGLYSERPHETY 907
2907 CTTTATGTCGAACCTCTGCGCTACCGCAGCGCAATTAAGCTGCGCG 2956
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
907 RYRLEUTHRLAPLEUSERASNLYSGNLGLYASPLYSVALVALTHRL 924
2957 AAGTTCGCAAGGCACTTACCTTGCTGCTGTCAACATACCGCAACGA 3006
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
924 YSERLATHRLGLYASNPHETHRLGULNALALASPLYSRTHRLGLY 940
3007 CCGGTAGCTCGAGCAATGACGAGTAGGAGGAAAAAGACACACACC 3056
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
941 PROTHRLYS...ASNLEUTHRLERHRLPHEASPLASERASALATHR 955
3057 GCTGTCGAAAAATCTTAATTTACCCCTGCAAAAAGACAGCTGATGCCG 3106
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
956 ...ARGASNASNLEUASNVALSERLEUVALGLYSNTHRLVALAPLEUG 971
3107 GCGCATGCGGTTATCAGCTTATCCGCAAGACGCGAGTTCGCTGCAT 3156
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
971 LYALATRPYSTYRYSLEUARGASNVALASNGLYARGYRASPLEUTYR 987
3157 AATCGG..... 3162
988 ASNPTRGLUVALGLULYSARGASNGLNTHRVALASPTHTHRSNILETH 1004
3162 3162

1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsnG 1021
3162 3162
1021 IuGIuIleAlaArgValGIuThrProValProProAlaProAlaThr 1037
3162 3162
1038 ProSerGIuThrThrGIuThrValAlaGluAsnSerLysGlnIuSerLy 1054
3163 GTCAAAGACAAAGAGCTTCCGACAAACTCGCAGACGCG 3202
1054 sThrValGIuLysAsnGlnIuAspAlaThrGIuThrAlaGlnAsnG 1071
3203 GAGA..... 3207
1071 IuGIuValAlaGluGluAlaLysProSerValLysAlaAsnThrGIuThr 1087
3208 ACAGA 3212
1088 AsnGluValAlaGlnSerGIuSerGIuThrGIuGlnThrGIuThrGI 1104
3213 GCGCGCTTGACGCA..... 3228
1104 uIleLysGIuThrAlaLysValGIuLysGIuLysAlaLysValGIuL 1121
3228 3228
1121 ySclGIuLysAlaLysValGIuLysAspGIuIleGlnGluAlaProGln 1137
3229 AAACAGCGCAACTTCCGCGCAAA..... 3252
1138 MetAlaSerGIuThrSerProLysGlnAlaLysProAlaProLysGIuVa 1154
3252 3252
1154 lSerThrAspThrLysValGIuGlnThrGlnValGlnAlaGlnProGln 1171
3252 3252
1171 hrGIuSerThrThrValAlaAlaGlnAlaThrSerProAsnSerLys 1187
3253 CAACAGCGGAAAAAAGACAAACGCGCAACCTTGA 3287
1188 ProAlaGluGIuThrGlnProSerGIuLysThrAsnAlaGluProValThr 1204
3288 CGCGCTGATTGCGCGCGCGCAAT..... 3312
1204 rProValIleSerLysAsnGlnThrGluAsnThrThrAspGlnProThrc 1221
3313 GCCACGGAAGAGCAGAAATGTTGCC 3339
1221 IuArgGIuLysThrAlaLysValGIuThrGIuLysThrGlnIuProPro 1237
3340 GAACCGCGCGCGAGCGAGC..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGlnSerGIuThrValGI 1254
3361 GGGAAATGCC...GGCATTATCGACGCG 3388
1254 nProGlnAlaValIleuGlnSerGluAsnValIProThrValAsnAlaG 1271
3389 AGGAAGAGAAAAAGGTCGACGCGATTAAGAC..... 3423
1271 IuGIuValGIuAlaGlnIleuGlnThrGlnThrSerAlaThrValIleThr 1287
3424 ACCGCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGIuSerAlaThrAlaI 1304
3432 GCGGAACACCGGAGCGGCGGCTACC..... 3468
..... 3468

1304 eThrGIuThrAlaGluLysSerAspLysProGlnThrGIuThrAlaIAs 1321
3468 3468
1321 eThrGluAspAlaSerGlnIleLysAlaAsnThrValAlaAspAsnSer 1337
3469 ACCGCTTCCCGCGCGCGCGCGCGCGCG 3500
1338 ValAlaAsnAsnSerGIuSerGIuProLysSerArgArg...ArgAr 1353
3501 GGATTTCGCCACCGCAG..... 3519
1353 gSerIleSerGlnProGlnGIuThrSerAlaGluGIuThrThrAlaIAs 1370
3520 CCCAACCGCAA 3531
1370 eThrAspGIuThrThrIleAlaAspAsnSerLysArgSerLysProAsn 1386
3532 CCCAACCGCAG..... 3543
1387 ArgArgSerArgArgSerValArgSerGIuProThrValThrAsnGlySe 1403
3544 CCGACCTGATCAGCGTTATGCCATA 3571
1403 rAspArgSerThrValAlaLeuArgAspLeuThrSerThrAsnThrAsnA 1420
3572 GCGGTTTGAGTGAATTTCCGCC..... ACCTCAACAGC 3606
1420 lValIleSerAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal 1436
3607 GTTTCGCGCTACAGCAATTGACCGCGTG...TTGGCGAAGACG 3653
1437 GlyLysAlaValSerGlnIleSerGlnIleuGluMetAsnAsnGIuGI 1453
3654 CCGCAACGCGCTTGAGCAACGCGCATCGGACACCAACACTACCGTT 3703
1453 yGIuTrAsnValTrpValSerAsnThrSerMetAsnGluAsnTrpYers 1470
3704 GCGAAGATTTCGCGCGCTACCGCCAAACCGACTCGCCAAATCGGT 3753
1470 eSerGIuTrArgArgPheSerSerLysSerThrGlnThrGlnIleuGI 1486
3754 ATGCAGAAAAACCTCGGACGCGG...CGCGCGCATCTGTTTCSGA 3800
1487 TrpAspGlnThrIleSerAsnAsnValGlnIleuGIuValPheThrTy 1503
3801 CAACCGGACCGGAACACCTTGACGACGCGATCGGCACTCGGACGCGC 3850
1503 rValArgAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuA 1520
3851 TTGCCACGCGTCCGTTTTCGGGCAATACGGCATCGGACGATTCGACATC 3900
1520 lAgIuValAsnPheTySerLysTyTyAlaAspAsnIleTrpTyLeu 1536
3901 GGCATCAGCGCGCGGCTTGTAGTCGCGCACTTTCAGACGCGCAT 3950
1537 GlyIleAspLeuGIuTyGIuLysPheGlnSerAsnLeuLysThrAsnHI 1553
3951 CAGAGCAAAATCCGCGCGCGCTGCTCATTTACGCGATTCAGGCAAGAT 4000
1553 sAsnAlaLysPheAlaArgIleThrAlaGlnPheGIuLeuThrAlaGI 1570
4001 ACCGCGCAGGTTTCGGGATTCGGCATCGCAACCGACATCGGCGCAACG 4050
1570 ySalPheAsnLeuGIuAsnPheGIuIleThrProIleValGIuValArg 1586
4051 CGCTATTTCGCCAAAAAGCGATTACGATACGAAAACTCAATATCGC 4100
1587 TySerTyLeuSerAsnAlaAsnPheAlaLeuAlaLysAspArgIleLy 1603
4101 CACCCGCGGCTTGATTCACACCGCTACCGCGCGGCTTAAGCAGAT 4150
1603 sValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuSer 1620

```

4151 ATTGATTCAAACCGCGCACACATTCATCCAGCCCTTATTGAGCCTG 4200
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1620 yrlthryhisleuglyglu...pheserValthrprolleuSerAla 1635
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
4201 TCCAT...ACCGATGCCGCTTCGCGCAAGTCCGAAAGCGCGCATAC 4247
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1636 ArgTyrAspThrAsnGlnGlySerGlyLysile.....AsnValAsnG1 1650
4248 CGCCGATATGGCGAGATTCGCGAAACCGGAGTGGGATGAGGCGG 4297
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1650 nTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1667
4298 TAAACGCGCAATCAAGGTTTCAAGCTGTCCTTCACAGCTCCGCGCC 4347
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1667 eulysleuLysTyrHisAsnValLysleuSerLeuIleGlyGlyLeuThr 1683
4348 AAGGCGCGCAATTGGAAGCGCGACAGCGCGCGCATCAATTAAGCTA 4397
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
1684 LysAlaTyrGlnAlaGlyLysGlnLysThrAlaGluLeuLysLeuSerPh 1700
4398 C 4398
1700 e 1700

```

seq_name: /cgn2_6/prodata/1/laa/PCTUS_COMB.pep:PCT-US95-10661A-5

```

seq_documentation_block:
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-Aug-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-Aug-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
PCT-US95-10661A-5

```

alignment_scores: Length: 1867
 Quality: 1581.50
 Ratio: 1.677
 Percent Similarity: 50.509 Percent Identity: 25.174

alignment_block:

us-09-303-518d-653 x PCT-US95-10661A-5 ..

Align seg 1/1 to: PCT-US95-10661A-5 from: 1 to: 1702

```

64 CGCTTCGCGCGCTTACCTTAGCCATATGCTGTGCTGCGCATTCGCGC 113
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
5 LysPheLysleuAsnPhelIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
114 CCAAGCGCGCGCGGACACACTATTTCGCGATCCACTACACTACTACTAC 163
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
21 oTyrThrGluAlaAlaLeuValAlaArgAspAspValAspIyrGlnIlePhe 38
164 GCGACTTTCGCGCAATTAAGCAAGTTCAGTCGCGCGCGCAAGATATT 213
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
38 rGAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
214 GAGCTTTACACAAAAAAGGGAGTGTGGCGCAATTCAGTACGAAAGC 263
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
55 GluValAlaArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG1 71
264 C...CCGATGATGATTTTCTGTGATCG...CGTAACGGCGTGGCGG 307
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaTr 88
308 CATTCGCGCGCGCATCATATATTGTAGCGTGCACAT...AAGCGCGC 354
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATACAAATGTTGATTTTGGT.....GCGAGCGAAGCAA 389
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAlaAsnGlyAs 121
390 TCCCGATCAGCACCGC.....TTTCTTACCAATG 421
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
121 nasPlySerHisArgAspValSerSerGlnGluAsnArgTyrPheSerV 138
422 TGAAGAATAATATTAAGACGAGGACTAACGC..... 456
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
138 alGluLysAsnGlnTyrProThrLysLeuAsnGlyLysAlaValThrThr 154
457 .....CATCTTATGGCGCGCATATCATATCCGCGTTCGCA 494
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171
495 CAATTGTGCACGATGCGACAGACCTGTGAG...ATGACCACTTATATG 541
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
171 pLysPheValThrGluValAlaIleProIleGluAlaSerThrAlaSerSera 188
542 ATGGGTGCAAAATACGCTGATTTAAATAATACCGTATGCTGTGCAATC 591
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
592 GGAGCAGCGACACATATTGGCGGCTGTGATGAAGCAAA..... 630
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuI1 221
631 .CCCAATAACCGGAGAAATTCATATCAT.....ATTGCAAGCG 667
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238
668 CATATTCCTGCTGCTGCGTGGCAATACCTTGCACAAAATGATGACGT 717
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
238 lArThrThrTyrGlyIleAlaGlyThrProTyrLysValAsnGluAsn 254
718 GGTGCGACAGTCACTTAGGTAGCGAAAAAATTAAACATAGC...CATA 764
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGlnGluHisSerAspProLy 271
765 TGGTTTTTACCAACAGGA.....GGCTCATTTGGCGACA 799
||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValIleuGlyAspS 288

```



```

800 GTGGCTACCAATGTTTATCTATGATGCCAAAGCAAAAGTGTAAAT 849
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
288 erGlySerProLeuPheValIyTrsParGlyLysTrPLeuPhe 304
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
850 AATGGGATTATGCAACAGCACCCCTATATAGAAAAGCAATGGCTT 899
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
305 LeuGlySerTrsPheTrpAlaGlyTrsAsnLysLysSerTrpGlnG 321
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
900 CCAGCTAGTCTGTAAGATTGTTCTATGATGAATCTTCTGTCAGATA 949
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
321 uTrpAsnLLeuTrLysProGluPheAlaLysTrhValLeuAspLysAsp 338
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
950 CCCATTTCAGTATCTACGAACCATCATCAAAATGGAAATACTTTTAAAC 999
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
338 hr..... 338
1000 GACATAATAATAGCGCAGCAAAAATGANGCCAAACATAAACATATTC 1049
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
339 .....AlaGlySerLeuIleGlySerAsnTrhGlnTrs 350
1050 TCTACTATATGATTAAACACGACCGTT.....CAATTGT 1087
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
350 nTrpAsnProThrGlyLysTrhSerValIleSerAsnGlySerLysTrL 367
1088 TTAATGTTTCTTATCCGACAGACAGCAAGAACCTGTTTATCATGCTGCA 1137
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGGTCACACAGTATTCACACCACAGCTGANTATAGAGAAATATTC 1187
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
378 .....ThraspSerLysLysAsnAsnHisGlyLysSerValTrh 390
1188 CTTTATTGACAAAGAAAGTGAAATTGATATCTACACAGCAACATCAACC 1237
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
390 rLeu.....ArgGlySerGlyTrhLeuTrhLeuAsnAsnHisLyspG 405
1238 AAGCGCGGGGGGTTTGTATTTTGAGGGTAAATTATTCGGTC...TCCCT 1284
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
405 InGlyAlaGlyGlyLeuPhePheGlnGlyLysPTrGlyValLysGlyTrh 421
1285 AAAACACACGAAACGTGGCAAGCGCGCGCTCATATCAGTATGTCAG 1334
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
422 SerAspSerTrhTrhTrpLysGlyAlaGlyValSerValAlaAspGlyL 438
1335 TACCGTACTTGGAAAGTAAACGGCTGGCAACGACCGCTGCCAAA 1384
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
438 sTrhValTrhTrpLysValHisAsnProLysSerAspArgLeuAlaLys 455
1385 TCGGCAAGCGACGCTGCTGCTCAACGCCAAAGGGGAAACCAAGGCTCG 1434
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
455 leGlyLysGlyTrhLeuIleValGlnGlyLysGlyLysAsnLysGlySer 471
1435 GTGCGCTGGCGCAGCGTAAGTCATCTTGATCAGAGCGGACATCA 1484
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
472 LeuLysValLyspGlyTrhValIleLeuLysGlnGlnAlaAspAlaAs 488
1485 AGGCAAAAACCAAGCCCTTAGTGAATCGCTTGTCAGCGGAGGAGGA 1534
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
488 asnLysValLysAlaPheSerGlnValIleValSerGlyArgSer 505
1535 CGGTGCAATGCAATGCCGATATCACTTCAACCCGACCAAACTCTATTTC 1584
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
505 hrValValLeuAsnAspLysGlnValAspProAsnSerLLeuTrpPhe 521
1585 GGGCTTCGCGCGGAGCTTGGATTGACAGCGGACATTCGTTTCGTTCCA 1634
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
522 GTrpPheTrpGlyGlyArgLeuAspAlaAsnGlyAsnAsnLeuTrhPheG 538
1635 CCGCATTCAAATACCGATGAGGGGCGATGTTGCAACCAACATCAAG 1684
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 uHisLLeuArgAsnLLeAspAspLysAlaArgLeuValAsnHisAsnTrh 555

```

```

1685 ACAAGATCCACCGTTTACCATTTACAGCAATTAAGATATTTACT..... 1728
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
555 euLysTrhSerTrhValTrhLLeuTrhGlyLysTrLeuLLeuTrhAspPro 571
1728 ..... 1728
572 AsnTrhLLeuTrhProTrpAsnLLeAspAlaProAspGluAspAsnPro 588
1728 ..... 1728
588 rAlaPheArgArgIleLysAspGlyGlyGlnLeuTrpLeuAsnLeuGluA 605
1728 ..... 1728
605 snTrpTrhTrpTrpAlaLeuArgLysGlyAlaSerTrhArgSerGluLeu 621
1729 .....ACAACCGCAATTAAC..... 1743
622 ProLysAsnSerGlyLysLysSerAsnGluAsnTrpLeuTrpMetGlyLysTrh 638
1744 .....AACAACTTGATACCAAAAAG 1765
638 rSerAspAlaAlaLysArgAsnValMetAsnHisLLeAsnAsnGluArg 655
1766 AATTCCTTACAAACGGTTGGCGAGAGAAGATGCAACCAACCAAGCAAC 1815
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
655 euAsnGlyPheAsnGlyTrpPheGlyGlnGlnGlnGlyLysAsnAsn 670
1816 GGGCGGCTCAATCTGAATTACCAACCGGAGAGAGCGGATCGCATTTACT 1865
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
671 GTrpAsnLeuAsnValTrhPheLysGlyLysSerGlnGlnAsnArgPheLe 687
1866 GCTTCCGCGGACCAATTTAAAGCGCAATATCAACGCAAAACAGGCA 1915
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
687 uLeuTrhArgGlyGlyTrhAsnLeuAsnGlyLysPLeuLysValGlnLysGly 704
1916 AACTGTTTTCAGCGCGCAGACCGACAGCGCGCTTACATATTTAGGA 1965
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
704 hrLeuPheLeuSerGlyArgProTrhProHisAlaArgAspLLeAlaGly 720
1966 AGCGGGGTCAAAAATGGAAGT.....ATCCACAGAGAAATCGT 2009
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
721 IleSerSerTrhLysLysAspGlnHisPheAlaGluAsnAsnGlnValVa 737
2010 GTGGGCAACGATTTGATGAGCCGACGATTTAAACGCAAACTTCATTA 2059
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaTrhAsnLLeAsnV 754
2060 TTCAGGCGCGACAGCGGTGTTTCC...CGCAATGTGGCCAAAGTGGA 2106
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
754 alTrhAsnAsnAlaTrhLeuTrpSerGlyArgAsnValAlaAsnLLeuTrh 770
2107 GCGGATTTGCGATTAAAGCAATCAGCCCAAGCGATTTTCGTGTCAGCC 2156
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
771 SerAsnLLeuTrhAlaSerAspAsnAlaLysValHisLLeGlyTrp..... 785
2157 GCATCAAAACCCACACATCTGTACACGTTTCGAGCTGGACGGGTTCGACA 2206
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
786 .LysAlaGlyAspTrhValCysValAlaArgSerAspTrpTrhGlyTrpAl 802
2207 GTTGTACCGAAAAAACCATTTACCGACGATTAAGTATGCTTCATTTAGC 2256
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
802 hrCysTrhTrhTrpLysLysLeuSer...AspLysAlaLeuAsnSerPheAsn 817
2257 AAGCCGACATCAGAGCAATGTCAGCTTGCAGATCAGCTCATTTTAA 2306
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
818 AlaTrhAsnValSerGlyAsnValAsnLeuSer..... 828
2307 TCTCACAGAGACTTCCACATCAACGCGCAATCTTAGTCAGCGGAGACA 2356
828 ..... 828
2357 CGCACTATACGGTTACGCGCAACGCCACCAAAAGCGCAACCTCAGCTC 2406

```

828 828
2407 GTGGGCAATGCCCAAGCAATTAAATCAAGCCATTAACGGCAAC 2456
|||||.....
829 ...GlysnAlasnPhleValLeuGlyLysAlaAsnLeuPhleGlyThrI 844
2457 ATGGCTTCGGACAATGCTTCATTTAATCTAAGCAACAGCCGTACAA 2506
|||||.....
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn..... 857
2507 ACGGCACTGTAGCGCTTCCGACAGCCTAAGCAACGTAAGCATTCC 2556
857 857
2557 GCACTCAAGCGCAATGTCTCCCTAGCCGATAGCAGTATTCATTGGA 2606
857 857
2607 AAACAGCCGCTTACCGGAAATACAGCGCGCAAGATACGGCATTAC 2656
857 857
2657 ACTTAAAGACAGCAATGACGCTGCCGTGGGACAGCAATTAGGCAT 2706
|||||.....
858SerHisThrPheLeuThrGlyAspSerAsnValAsnGln 870
2707 TTAACCTTGACAAGCCACCATTACACTCAATTCGCGCTATGCAcCGA 2756
|||||.....
871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnsAsn 885
2757 TGGCGCAGCGCGCAACGGCGAGTGGCGCAGATGCCCGCGCGCGCT 2806
|||||.....
885 pAlaAsnLysValThrThr..... 891
2807 CGCGCGCTCCCTATTCGCTTACGCCCAACTTCGCGCAGATCCCT 2856
891 891
2857 TTCAACACGCTGACGCTAAACGGCAATTGACGCTAGGACATTCGG 2906
|||||.....
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTy 907
2907 CTTATGCGGAACCTTCGCGCTACCGCAGCGCAATGCAATGAGTGGCG 2956
|||||.....
907 rTyLeuThrAspLeuSerAsnLysGlnGlyAspLysValValThrL 924
2957 AAAGTCCGAAGGCATTACACTTGGCTGTCAACAATACCGGCAACGAA 3006
|||||.....
924 ySerAlaThrGlyAsnPherLeuGlnValAlaAspLysThrGlyGlu 940
3007 CCGGTAGTCTCGAGCAATTGAGGTAGTGGAGAGAAAGCAACACACACC 3056
|||||.....
941 ProThrLys...AsnGluLeuThrLeuPhenSerAlaSerAsnAlaThr.. 955
3057 GCTGTCCGAAATCTTAATTCACCTGCAAAAGCAACAGCTGCATCCG 3106
|||||.....
956 ...ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAsnLeu 971
3107 GCGCATGGCGTTATCAGCTTATCCGCAAGACGGCGAGTTCGCGTGCAT 3156
|||||.....
971 LysAlaTrpLysTyLysLeuArgAsnValAsnGlyArgTyAspLeuTy 987
3157 AATCG..... 3162
|||||.....
988 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsn 1021
3162 3162

1021 LuGluIleAlaArgValGluThrProValProProAlaProAlaThr 1037
3162 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnGlySer 1054
3163GTCAGAAGAACAGACTTCCGACAAACTCGCAAGCGG 3202
|||||.....
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsn 1071
3203 GAGAA..... 3207
|||||.....
1071 LysGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208ACGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrGlu 1104
3213 GCGCGCTTGACGCA..... 3228
|||||.....
1104 uLleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGlu 1121
3228 3228
1121 ySerGluLysAlaLysValGluLysAspGluIleGlnLysAlaProGln 1137
3229AACAGGACACACTTGGCGGCAAA..... 3252
|||||.....
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVal 1154
3252 3252
1154 lSerThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGln 1171
3252 3252
1171 hrGlnSerThrThrValAlaAlaGluAlaThrSerProAsnSerLys 1187
3253CAACAGCGGAGAAAGACAGCGCAAGGCTTGA 3287
|||||.....
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204
3288 GCGCGTGAATGGCGCGCGCGCAAT..... 3312
|||||.....
1204 rProValAlaSerLysAsnGlnThrGluAsnThrThrAspGlnProThr 1221
3313GCCACGAAAGGCAAGAGTTGCGC 3339
|||||.....
1221 LuArgGluLysThrAlaLysValGluThrGluLysThrGlnLeuProPro 1237
3340 GAACGCGCGCGGAGGAGGAGC..... 3360
|||||.....
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValGlu 1254
3361GGGAAAAATGCC...GGCATTAATGACGCGG 3388
|||||.....
1254 nProGlnAlaValLeuGluSerGluAsnValProThrValAsnAlaG 1271
3389 AGCAAGAGAAAAACGGGTGACGCGGATPAAGAC..... 3423
|||||.....
1271 LuGluValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424ACGCGCTT 3431
|||||.....
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaI 1304
3432 GCGGAACAGCGGAAGCGGAACCGCGCGGCTAC..... 3468
|||||.....
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAla 1321
3468 3468
1321 eThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337

[illegible]

```

Alignment_scores:
  Quality: 1515.00      Length: 1725
  Ratio: 1.733          Gaps: 44
  Percent Similarity: 50.667      Percent Identity: 26.783

Alignment_block:
US-09-303-518D-653 x 5268270-2  ..

Align seg 1/1  to: 5268270-2  from: 1  to: 1507

49  AAACCGCGCGCATCCGGCTTCTCGCCCGCTTACTTACCATATACCTGTC  98
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
2  LysAlaIysArgPheIysIleAsnAlaIleSerLeuSerIlePheLeuAl  18
49  GTTCGGCACTTGTGCCCCAACGCCGGCGGACACACTTATTTGGCATCA  148
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
18  aTygAlaLeuThrProTyrSerGlnAlaIleValaIArgAspVala  35
149  ACTACCACTACTACCGGCACTTCCGCAAAATAAGCAGCTTGCAGTC  198
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
35  sPTyGlnIlePheArgAspPheAlaGlnAsnIysGlyIysPhePheVal  51
199  GGGCGGAAGATATTATGGATTACAAAAAAGGGAGATGGTCGGCAA  248
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
52  GAlaAlaThrAspLeuSerValIysAsnIysArgGlyGlnAsnIleGlyAs  68
249  ATCGATGACGAAGACCCCGATGATGATTTTGTGTGTGATCCGCTAAC  297
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
68  nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL  85
298  ..GGCGTGGCGGCACTTGGCGGGCGCATATATATTGTGACCGTGCACAT  345
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
85  ySArgIleAlaThrValValaIAspProGlnTyrAlaValaSerValIysHis  101
346  .....AACGGCGCATATACATGTGTA  368
102  AlAlaIysAlaGlnValaHisThrPheTyrTyrGlyGlnTyrAsnIleHis  118

```


647
utrrlglyalaasnlleuansnglylsllleservallthcldnlysnv 664
1919 tgrttttcagccggcagaccgacacccgaccccttacatcatTTTAGGAAC 1368
.....
664 alleuenserlglyargProthrProHISAlaArgAspPheValAsnLys 680
.....
1969 ggggtgctAAAAATGGAAGT.....ATCCCAAGAGGAATCGGTG 2012
.....
681 SerSerAlaArgLysAspAlaHisPheSerLysAsnAsnGluValIlePh 697
.....
2013 GCAACAAGATTGATGACCGACCATTTTAAAGCGAAACTTCATATT. 2061
.....
697 egluAspArpTrpIleAsnArgThrPheLysAlaIleaglulIeAlaVal 714
.....
2062 ..CAGGGCGACAAGGGGTTCGCCCATGTTCGCATGTTCGCAAGTGAAGC 2109
.....
714 smglSerLasePheSerSerGlyArgAsnValAspAspIleThrAla 730
.....
2110 GATTGGCATTTTAGCAATCACCGCCCAAGCATTTTCGCTGCGACCGCA 2159
.....
731 AsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTr.....Ly 745
.....
2160 TCAAAGCCACAAATCTGTACACGTTGGACTGAGCGGGTGTGCAAGTT 2209
.....
745 sasnlglyAspGluValLysValArgSerAspTrpThrGlyTrValThrC 762
.....
2210 GTACCAAAAAACCAATACCGCATTAAGTATGGTTCATTCATGAGAAC 2259
.....
762 yAsnThrClysnLeuSerAspLysAlaLeuAsnSerPheAsp**Ala 778
.....
2260 ACCGACATCAGAGCAATGTACGCTTCGCGATCAGCCGTCATTTAAATCT 2309
.....
779 ThrLysIleAsnGlyAsnValAsnLeuAsnGlnAsnAlaIleValLe 795
.....
2310 CACGAGACTGGCCACATCAAGCGAATCTTAAGTGCAGGGGAGACACGC 2359
.....
795 u.....GlyLysAlaIleAlaLeuTrpGlyLysIleGlnGly..... 806
.....
2360 ACTATACGGTTACGGCAACGCCAACCCAAACGGCAACTCAGCCCTGTG 2409
.....
807Gln 807
.....
2410 GGCATGGCCCAAGCAACATTTATCAAGCCCATTTAAAGCGACACATC 2459
.....
808 GlysnSerArgValSerLeuAsnGln..... 816
.....
2460 GCGTTGCAACAATGCTTCAATTAATCAAGCAACAGCGCGTACAAAGC 2509
.....
816A... 816
.....
2510 GCAGCTGACGCTTTCGACACAGCGTAAGGCAAGCATGAACCATTCGCGCA 2559
.....
816 816
.....
2560 CTCACAGCAATGTCTCCCTAGCGATTAAGGACAGTATTCATTTGAAAA 2609
.....
816 816
.....
2610 CAGCGCGTTTACCGGAAAAATCAGCGCGGCAAGGATACGCAATTACACT 2659
.....
816 816
.....
2660 TAAAGACAGGAATGACGCTGCGGTCGGGCAACGGAATTAAGCATTTA 2709
.....
817HisSerLysTrpHisLeuThrGlyAspSerGlnValHisAsnIle 831
.....
2710 AACCTTGAACAAGCCACCATTAACATCAATTCGCGCTATGACACAGATCG 2759
.....
832 SerLeuAlaAspSerHisIleHisLeuAsnAla.....SerAspI 846
.....
2760 GGCAGGCGCGCAACCGGCAAGTGGCGAGATGGCGCGCGCGCGTTGCG 2809
.....

```

3516 ..... 3516
1119 laargvalprolleargmetglarggluleuleasnasptyrglglu 1135
3516 ..... 3516
1136 lleproleuglululegluasplualaglugluluararglinal 1152
3516 ..... 3516
1152 athrlnpheniserryserargasnarargalalleseerclup 1169
3516 ..... 3516
1169 roseraseraspluasplaserluservalserthreraplyshis 1185
3516 ..... 3516
1186 proglinspasnthrgluleuhisglulysvalgluthralaglyleugl 1202
3517 ..... CAGCCCAACCCGCAACCCGACGCGACCTGA 3553
1202 nproarglalaaglproargthrglnalalaalaglnalasplav 1219
3554 TCAGCCGTTATGCCAATAGCGGTTGAGTAATTTCCGCCACGCTGAC 3603
1219 alserthrasnthrasneralaleuserasplamealaserthrgln 1235
3604 AGCGT.....TTCGCCGTACAGAGCAATTTGACCG 3635
1236 Serlleuleuaspthrglyalatylerthraghislleaglinal 1252
3636 CGTGTTCGCCAAGACCCGCCGCTTTGGACAAGCGGCAATCCGGG 3685
1252 sserarglalsplaglulysasnservaltrpmeteluasnthrglyt 1269
3686 ACACCAACATACCGCTGCCAATTTCCGCCGTACCGCACACAAC 3735
1269 yrglyargasptylalaserlalntrargalgheserlyarg 1285
3736 GACCTGCCCAATCGGTATGCAGAAAAACCTGCCAGCGG...CGCGT 3782
1286 Thrlnthrglnlleglylleaspargserleusergluasmetglnl 1302
3783 CGGCATCTGTTTCGCACACCGGACCGGAACACCTTCAGACGCGCA 3832
1302 egllyvalleuthrlyrseraspserglnhsthreaspelnala. 1318
3833 TCGGCACTCGGCACGCTGCCACAGGCTTTTCGGGCATATACGCG 3882
1319 ..gllyglylvasnthrphelvalglnalasnleuthrglylytyr 1334
3883 ATCGGC...AGTTTCGACATCGGCATCGCGGCGGCTTTTAGTAG 3929
1335 leuasnasplalatrptyrvalalaglyasplleaglylaserleuar 1351
3930 CGGCAGCTTTCAGACGCGCATCAGAGCAAAATCCGCCGCTGCTCC 3979
1351 gserargleuglnthrglnlglulysalasnphasnargthserlleag 1368
3980 ATTACGCAATTCAGGACAGATACGCGCAGGTTTCGGGCAATTCGCG 4029
1368 lnthrglyleuthrleugllyasnthrlleuslileasnlnphelglnle 1384
4030 GAACCGCATCGCGCAACGCGCTATTTCTCCAAAAAGGATACG 4079
1385 ValProseralaglylleargtyrserargleuseralaaasptyrly 1401
4080 ATAGCAAAAGCAATATCGCCACCGCGCTTCATTCACACGCTAAC 4129
1401 sleugllyaspsaservalysValSerSerMetAlaVallysthrleut 1418

```

```

4130 GCGCGGCATTAAGCAGATATTCATTCAAACCGCGCAACATTTCC 4179
1418 hrleaglyleuasphelalatrargphelys...Valglylvasnthr 1433
4180 APCAGCCTTAATTTGACCTGTCTATACCATGCCGCTTCGCGCAAGT 4229
1434 VallysProleuuseralalatrype...AlaasnTyrglylysgl 1449
4230 CGGACGCGCGCAATACCCCGCTATTTGCCGCGCATTTTCGCAAAACC 4279
1449 ygllyalasnvalglylylserphenalatrlysalaspsanglng 1466
4280 GCAGTCGGCAATGGGCGTAAACGCCGAATCAAGTTTCACGCTGTC 4329
1466 lnglntrseralaglyvalalaleuleuthrglarsnvalthreusn 1482
4330 CTCACGCTGCCCGCGCAAGGCGCGCATTTGGAAGCGCACACAGCG 4379
1483 Valasnlyserlthrlysglylsglnleugllyllysserl 1499
4380 GGGCATCAATTAAGCTACCGCTGG 4404
1499 yglntlelysllelnglnleargphe 1507

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB pep: US-08-296-791-6

seq_documentation_block:

Sequence 6, Application US/08296791

Patent No. 6245337

GENERAL INFORMATION:

APPLICANT: St. Geme III, Joseph W.

TITLE OF INVENTION: Haemophilus Adherence and Penetration

TITLE OF INVENTION: Protein

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791

FILING DATE: 25-AUG-1994

CLASSIFICATION: A35

ATTORNEY/AGENT INFORMATION:

NAME: Treccartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1848 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-296-791-6

alignment_scores:

Quality: 1513.00

Ratio: 1.519

Percent Similarity: 51.983

Length: 1916

Gaps: 60

Percent Identity: 25.418

alignment_block:

US-09-303-518D-653 x US-08-296-791-6 ..

Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

```
64 CGCTTCGCGCGGCTTACTTACCATATGCGTCTGCTTCGCGCATTTGCGC 113
   :::::::::::::::::::::
5  LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
   :::::::::::::::::::::
114 CCAAGCCGCGGCGGACACACTTATTTGCGCATCAGTACCAATACATATC 163
   | :::::::::::::::::::::
21 oTyrThrGlnAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePheA 38
   :::::::::::::::::::::
164 CGCATTTGCCGAATAAAGCAAGTTTGCGAGTGGCGGCGCAAGATATT 213
   ||:::::::::::::::::::
38 rGaSPheAlaGlnAsnLysGlnLysPheSerValGlnAlaThrAsnVal 54
   :::::::::::::::::::::
214 GAGGTTCACAAACAAAAGGAGTGTGTGCGCAATCGATGACGAAGC 263
   ||:::::::::::::::::::
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGln 71
   :::::::::::::::::::::
264 C...CCGATCATGTGATTTTCTGTGTATCG...CGTACGGCGCGTGGCGG 307
   : ||:::::::::::::::::::
71 yIlePheMetIleAspPheSerValAlaAspValAspLysArgIleAlaTr 88
   :::::::::::::::::::::
308 CATTGGCGGCGGATCAATATATGTGTAGCGTGCACAT...AACGGCGGC 354
   :::::::::::::::::::::
88 hrLeuValAsnProGlnTyrValAlaGlnValLysHisValSerAsnGln 104
   :::::::::::::::::::::
355 TATAACAATGTGATTTTGGT.....GCGGAGGGAAGCAA 389
   :::::::::::::::::::::
105 ValSerGlnLeuHisPheGlnLysLeuAsnGlnLysAsnMetAsnAsnGln 121
   :::::::::::::::::::::
390 TCCCGATGACGACCGC.....TTTCTTACCAAAATTG 421
   | :::::::::::::::::::::
121 nAlaLysSerHisArgAspValSerSerGlnLysAsnArgTyrTyrThrV 138
   :::::::::::::::::::::
422 TGAACAAATTAATTAT.....AAGCA 444
   ||:::::::::::::::::::
138 aLglnLysAsnAsnPheProThrGlnAsnValThrSerPheThrLysGln 154
   :::::::::::::::::::::
445 GGCACATACGCGCATCCTTATGCGCGATATATCATATGCGCGGTTGCA 494
   :::::::::::::::::::::
155 GlnGlnAspAlaGlnLysArgArgGlnAspTyrTyrMetProAlaGlnAs 171
   :::::::::::::::::::::
495 CAAATTTGTCACAGATGACAGAACCTGTTGAG...ATGACAGTATATATG 541
   ||:::::::::::::::::::
171 pLysPheValThrGlnValAlaProIleGlnAlaSerThrAlaAsnAsnA 188
   :::::::::::::::::::::
542 ATGGGTGGAATACGCTGATTAAATAAATACCTGATCGTGTTCGATC 591
   :::::::::::::::::::::
188 snLysGlnGlnTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204
   :::::::::::::::::::::
592 GAGACGACGACACAAATAT.....TG 611
   ||:::::::::::::::::::
205 GlnSerGlnThrGlnPheIleTyrLysGlnSerArgTyrGlnLeuI 221
   :::::::::::::::::::::
612 CGCGTCTGATGAAGACGACCCATAAACCGGAAGTTCATATCAT... 657
   :::::::::::::::::::::
221 eLeuThrGlnLysAspLysGlnGlnLysLeuAlaGlnAsnTrpAspValG 238
   :::::::::::::::::::::
658 .....ATGGCAAGCGCATATCTTGGCTGCGTCGCGT 687
   :::::::::::::::::::::
238 LysLysAsnLeuGlnLeuValAlaLysAsnAlaTyrThrTyrGlnIleAla 254
   :::::::::::::::::::::
688 GGCATATCCTTTGGCACAATAATGATGATGCTGCGACAGTCAACTAGG 737
   ||:::::::::::::::::::
255 GlnTyrGlnTyrLysValAsnHisGlnAsnAsnGlnLeuIleGlnPheGln 271
   :::::::::::::::::::::
738 TAGGCAAAAATTAACATAGC...CCATATGCTTTTAAACAAACAGGA 783
   :::::::::::::::::::::
271 yAsnSerLysGlnGlnLysSerAspProLysGlnIleLeuSerGlnAspR 288
   :::::::::::::::::::::
784 .....GGCTCATTTGGCGACAGTGGCTCACCAATGTTATC 819
   :::::::::::::::::::::
   :::::::::::::::::::::
288 rOleuThrAsnTyrAlaValLeuGlnLysSerGlnYserProLeuPheVal 304
   :::::::::::::::::::::
820 TATGATGCCCAAAAGCAAAAGTGTAAATTAATGGGATTTGCCAAACAG 869
   ||:::::::::::::::::::
305 TyrAspArgGlnLysGlnLysTrpLeuPheLeuGlnSerTyrAspPheTr 321
   :::::::::::::::::::::
870 CAACCCCTATATGCAAAAAGCAATGGCTTCAGCTAGTTCGTAAGAT 919
   ||:::::::::::::::::::
321 pAlaGlnTyr.....AsnLysLysSerT 329
   :::::::::::::::::::::
920 GGTTCATATGATGAATCTTTGGTGGAGATACCATTCAGTATTCACGAA 969
   ||:::::::::::::::::::
329 rPheGlnLysTrpAsnIleTyrLysHisGlnPheAlaGlnLysIleTyrGln 345
   :::::::::::::::::::::
970 CCACATCAAAATGGCAATACTTTTAAACACAAATTAATATGCGGACG 1019
   :::::::::::::::::::::
346 GlnTyr.....SerAlaGln 350
   :::::::::::::::::::::
1020 AAAAATCGATGCCAAACATAAACATATCTCTACCTTATACATTAATA 1069
   :::::::::::::::::::::
350 ySerLeuIleGlnYserAsnThrGlnTyrThrPheGlnAlaThrGlnYserT 367
   :::::::::::::::::::::
1070 CACGAACCGTT.....CAATTGTTAATGTTCTTATACGAG 1107
   ||:::::::::::::::::::
367 hrSerThrIleThrGlnLysGlnLysProLeuSerValAsnLeuThrAsp 383
   :::::::::::::::::::::
1108 ACAGCAGAGAACCTGTATATCATCTGCAGGTGGGGGTACACATTAATCG 1157
   :::::::::::::::::::::
384 GlnLysAspLysPro..... 388
   :::::::::::::::::::::
1158 ACCCAGACTGAATTAATGAGAAATATTTCTTTATGCAAGGAAGAA 1207
   ||:::::::::::::::::::
389 .....AsnHisGlnLysSerIleThrLeu....LysGlnYserG 400
   :::::::::::::::::::::
1208 GTCAATTGATCTTACACAAACATCAACCAAGCGCGCGGTTGAT 1257
   ||:::::::::::::::::::
400 lYThrLeuThrLeuAsnAsnHisIleAspGlnGlnValGlnLysLeuPhe 416
   :::::::::::::::::::::
1258 TTTGAGCGTAATTTTACGGTC...TGCCTAAACAAACGAACCGTGGCA 1304
   ||:::::::::::::::::::
417 PheGlnGlnLysPyrGlnValLysGlnThrSerAspSerThrThrTyrLys 433
   :::::::::::::::::::::
1305 AGCGCGGCGGCTTCATATCATGATGATGCGATACCTTACTTGAAGTAA 1354
   ||:::::::::::::::::::
433 sGlnAlaGlnLysSerValAlaAspGlnLysThrValThrTrpLysValH 450
   :::::::::::::::::::::
1355 ACGGCGTGGCAACGACCGCTGTCCAAATGCGCAAGACGCTGCTG 1404
   :::::::::::::::::::::
450 lAsnProLysTyrAspArgLeuAlaLysIleGlnLysGlnYThrLeuVal 466
   :::::::::::::::::::::
1405 GTTCAAGCCAAAGGGAACCAAGGCTGCGTACAGTGGCGGAGCGGTAA 1454
   ||:::::::::::::::::::
467 ValGlnGlnLysGlnLysAsnGlnGlnLeuLeuLysValGlnLysPhe 483
   :::::::::::::::::::::
1455 AGTCATCTTAGATCAGCAGCGGAGATCAAGCAAGCAAAACAGCCTTAA 1504
   ||:::::::::::::::::::
483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500
   :::::::::::::::::::::
1505 GTCAATTCGCGTGTGTCAGCGGCGAGGGGACGGTCAACTGATCCGAT 1554
   ||:::::::::::::::::::
500 eGlnValAlaGlnLysLeuIleArgSerThrLeuValLeuAsnAspAsp 516
   :::::::::::::::::::::
1555 AATCAGTTCAACCCGACCAACTATATTTGGCTTTCGCGGCGGACGTTT 1604
   ||:::::::::::::::::::
517 LysGlnValAspProAsnSerIleTyrPheGlnPheAlaGlnLysValLe 533
   :::::::::::::::::::::
1605 GGAATTTGAACGGGCAATGCGTTTCTTCACCGCATCAAAATACGATG 1654
   ||:::::::::::::::::::
533 uAspLeuAsnGlnLysAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
   :::::::::::::::::::::
1655 AAGGGCGGATGATGTGTCAACCAATCAAGACAAAGAAATCCACGTTAC 1704
   :::::::::::::::::::::
   :::::::::::::::::::::
```



```
550 spgIyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATPACAGCATTAAGATATTACTACA.....ACCGCAATAA 1742
|||||..... |||||
567 IleThrGlySerLeuIleThrAsnProAsnThrIleThrSerTyAs 563
1743 C..... 1743
|
583 nIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIlePro 600
1743 ..... 1743
600 yTrArgGlnLeuTyRheAsnGlnAspAsnArgSerTyTyThrLeuLys 616
1743 ..... 1743
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAs 633
1743 ..... 1743
633 nGluAsnTrpLeuTyMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1744 .....AACACTTGATAGCAAAAAAGAAATGGCTTCAACGGTTGGTTT 1788
|||||..... |||||
650 alMetAsnHisIleAsnAsnGlnIleArgMetAsnGlyPheAsnGlyTyRhe 666
1789 GGGCAAGAAAGTGCACCAAAACG...AACGGGGGCTCAATCGAATTA 1835
|||||..... |||||
667 GlyGluGluGlnTrpLysAlaThrGlnAsnGlyLysLeuAsnValTrpH 683
683 eaSnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsnL 700
1886 TAAACGGCAATATCAGCGCAAAACGCAACGTTTTCAGCGGCGAGA 1935
|||||..... |||||
700 euAsnGlyAspLeuAsnValGlyLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCGACACGGCAGCGCTCAATCATTTAGGAAGCGGGTGGTCAAAATGGA 1985
|||||..... |||||
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 AGGT.....ATCCACAGAGAGAATCGTGGGACACAGATGATCG 2029
|||||..... |||||
733 pProHisPheThrGlnAsnAsnGlnValValValGluAspAspTrpIleA 750
2030 ACCGCACATTTAAAGCGGAAACTTCATATTCAGGGCGGCAACCGGTG 2079
|||||..... |||||
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnAlaSerLeu 766
2080 GTTTC...CGCAATGTTGCCAAGTGAAGCGGATGGCATTTAAGCAA 2126
|||||..... |||||
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAGTTTGGGTGTCGACCGCATCAAGCCACACAATCT 2176
|||||..... |||||
783 nasAlaGlnValHisIleGlyTyR...LysThrGlyAspThrValC 798
2177 GTACAGCTGGAGTGGACGGGCTGTGACAAAGTTGTACCGAAACAAAT 2226
|||||..... |||||
798 ySValArgSerAspTrpThrGlyTyRValTrpCysHisAsnSerAsnLeu 814
2227 ACCGACGATAAGTGAATGCTTCATTTAGCAAGACGACATCAGAGCAA 2276
|||||..... |||||
815 Ser...GluTySAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAs 830
2277 TGTACGCTTCCGATCAGCTCATTTAAATCTCAGAGACTTCCACAC 2336
|||||..... |||||
830 nValAsnLeuThrGlnAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846
2327 TCAACGGCAATCTTAGTCAGCGGA..... 2352
|||..... |||
846 eupheGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGln 862
2353 GACACGACTATACGTTACGGCAACGCCAC... 2385
|||||..... |||||
863 AsnSerHisTrpHisLeuThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
2386 .CAAAACGGCAACCTCAGCCTGTGGCAATGCCAA..... 2421
|||||..... |||||
879 uThrAsnGlnHisIleHisLeu....AsnAlaGlnAsnAspAlaAsnL 894
2422 ....GCACATTTAATCAGCCACATTAAGCGACACATCGGCTCG 2466
|||||..... |||||
894 ySValThrThrTyRAsnThrIleThrValAsn.....SerLeuSer 907
2467 GACAAATGCTTCATT.....AATCTAAGCAA 2492
|||||..... |||||
908 GlyAsnGlySerPheTyRTrpValAspPheThrAsnAsnLysSerAs 924
2493 CAACGCCGTACAAACCGCAGT..... 2514
|||||..... |||||
924 nLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValA 941
2515 .....CTGACGCTTTCGACAAAC 2532
941 laAspLysThrGlyGluProAsnHisAsnGlnLeuThrLeuPheAspAla 957
2533 GCTAAGCA.....AACGTAAGCATTCGCGCATCAACGGCAATGT 2573
|||||..... |||||
958 SerAsnAlaThrArgAsnAsnLeuGlnValThrLeuAlaAsnGlySerVa 974
2574 CTCCCTAGCCGATTAAGCGAGTATTCATTTGAAACACCGCCTTACCG 2623
|||||..... |||||
974 L.....AspArgGlyAlaTrpLysTyRAsnLysLeuArgAsnValAsnG 988
2624 GAAAAATCAGCGCGGCGGCAAGATACGCATTCACCTTAAAGACAGCGAA 2673
|||||..... |||||
988 LArgTyRAsp.....LeuTyRAsnProGlu 996
2674 TGGACGCTCCGTCGGGACAGGAATTAAGCAATTAACCTTGACACAGC 2723
|||||..... |||||
997 .....ValGluLysArgAsnGlnThrValAspThrTh 1007
2724 CACCATTAACATCAATTCGCGCTATCGACACAGATCGCGCGGCAAA 2773
|||||..... |||||
1007 rAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGln 1024
2774 CCGGCAGT..... 2781
1024 eTrAsnAsnGluGluIleAlaArgValGluThrProValProProAla 1040
2782 .....GCGGCAGATGGCGCGCGCGCGCT... 2805
1041 ProAlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAl 1057
2805 ..... 2805
1057 aGluThrAlaGlnProAlaMetCgluGluThrAsnThrAlaAsnSerTrg 1074
2806 .....TCGGGC 2811
1074 LuThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSer 1090
2812 CGTCCCTATTAATCGGTACG..... 2832
|||||..... |||||
1091 GluSerValProSerGluThrThrGluLysValAlaGluAsnProProG 1107
2832 ..... 2832
1107 nGluAsnGluThrValAlaLysAsnGluGlnGluAlaThrGluProTrp 1124
2833 .....CCGCCAATTCGGCGAGAA 2850
|||||..... |||||
1124 roGlnAsnGlyGluValAlaLysGluAspGlnProThrValGluLalaSn 1140
```

```
2851 TCCGCTTTCAACACGCTGACGTAAC...GCATAATTGAACGCTCAGG 2897
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1141 ThrGlnThrAsnGlnuAlaThrGlnSerGlnGlySerThrGlnGluThrGln 1157
2898 AACATTCGCGCTTTATGTCGAA... 2919
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1157 nThrAlaGlnThrLysSerGlnProThrGlnSerValThrValSerGluA 1174
2920 ..... 2928
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1174 snGlnProGlnLysThrValSerGlnSerThrGluAspLysValVal 1190
2929 TACCGCAGCGGCAATTCGAAGCTG... 2952
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1191 GluLysGlnGluLysAlaLysValGluThrGlnGluThrGlnLysAlaPr 1207
2952 ..... 2952
1207 ogLlnValThrSerLysGlnProLysGlnAlaGluProAlaProGluG 1224
2953 ..... 2973
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1224 LuValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnThr 1240
2974 TACACCTTGCGCTGC... 2988
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1241 GlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLysPr 1257
2988 ..... 2988
1257 oAlaGluGluThrGlnProSerGlnLysThrAsnAlaGluProValT 1274
2989 ..... 3024
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1274 hrProValValSerGlnAsnThrAlaThrGlnProThrGlnThrGlnGlu 1290
3025 TTGACGCTAGTGAGAGAAAGACACACACCGCTGCCGAAATCTTAA 3074
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1291 ThrAlaLysValGlnLysGlnLysThrGlnGluValProGlnVal... 1305
3075 TTTACCCCTGCAAGACACACGCTGCATGCCGGCGCATGGCGTTATACG 3124
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1306 ..... 1318
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13125 TTATCCGCAAGACGCGAGCTCCGCTGCATATCCGTCGTAAGACAA 3174
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1318 alYsPProGlnAlaGlnThrLysProGlnAlaGluProAlaArgLysAsn 1334
3175 GAGCTTCCGACAAA... 3189
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1335 ValLeuThrThrLysAsnValGlyGluProGlnProGlnAlaGlnProGln 1351
3190 ..... 3220
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1351 nThrGlnSerThrAlaValProThrThrGlnGluThrAlaAlaAsnSerL 1368
3221 TGACGGCAAAAACAGCACACTTCCGCCAAACAGCGCGAA... 3264
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1368 ySPProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAla 1384
3265 ..... 3287
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1385 ArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrse 1401
3288 CGCGCTATATGGCGCGCGGCAATGCCAGGAAAGGCAAGAGTGTG 3337
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1401 rAlaThrValSerThrGlnProAlaLysGluThrSerSerAsnValG 1418
3338 CCGAAGCGGCC... 3348
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1418 LuGlnProAlaProGluAsnSerLysnThrGlnLysSerAlaThrThrMet 1434
3348 ..... 3348
```

```
1435 ThrGlnThrAlaGlnLysSerAspLysProGlnMetGluThrValThrGln 1451
3349 .....CGGCAG...GCAGCGGGGAAATGCCG 3373
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1451 uAsnAspArgGlnProGlnAlaAsnThrValAlaAspAsnSerValAla 1468
3374 GCATTTATGCAGCGCGGAGAGAGAAAAACGG...GTGCAG 3411
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1468 snAsnSerGlnSerSerGlnSerLysSerArgArgArgSerValSer 1484
3412 GCGGATAAAGACACCGCTGGCGAAACAGCGGAAACCGGCGGCG 3461
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1485 GlnProLysGlnThrSerAlaGlnGluThrThrValAlaSerThrGlnGln 1501
3462 GCGTACC...ACGCGCTTCCCGCGCGCGCGCGG 3493
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1501 uThrThrValAspAsnSerValSerThrProLysProAspSerArgArgT 1518
3494 CCGCGCGGATTTGCCGCAACCGCACCCCAACCGCACCCACCGCGAG 3543
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1518 hrArgArgSerValGlnThrAsnSerTyrgLuproValGluLeuProThr 1534
3543 ..... 3543
1535 GluAsnAlaGluAsnAlaGluAsnValGlnSerGlnAsnAsnValAlaAs 1551
3544 .....CGGACCTGATCAGCGTTATGCCAATACCGGTT 3577
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1551 nSerGlnProAlaLeuAlaGlnLeuThrSerLysAsnThrAsnAlaVal 1568
3578 TGAGTGAATTTCCGCC...ACGCTCACAGCGCTTTC 3612
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1568 LeSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLys 1584
3613 GCGGTACAGAGCAATTTGACCGCGTG...TTGCCGAGACCGCGCAA 3659
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1585 AlaValSerGlnHisLeSerGlnLeuGluMetAsnGlnGlyGlnTy 1601
3660 CCGCGTTTGACAGCGGCGATCCGGGACACCAACACTACGCTTGCAAG 3709
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1601 rAsnValTrpIleSerAsnThrSerMetAsnLysnTyrsSerSerGluG 1618
3710 ATTTCGCGCCTTACCGGCAACGACGCTGCGCAATTCGTTGACG 3759
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1618 InTyArgArgPheSerSerLysThrGlnThrGlnLeuLysTyrrpSP 1634
3760 AAAAACCTCGGACGCGG...CGGTCGCGCATCTGTTTCGACACACG 3806
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1635 GlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTyValAr 1651
3807 GACCGAACAACCTTCGACGACGCGCATCGGCACTCGGACGCGTTGCC 3856
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1651 gAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaGlnV 1668
3857 ACGGTCGCTTTTCGCGCAATACGCGATCGGAGGTGACATCGCGCATC 3906
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1668 alAsnThrLysSerLysTyrrAlaAspAsnHisTyrrTyLeuGlyLe 1684
3907 AGCGCGGCGCGGCTTTAGTAGCGGACCTTTGACAGCGCATCAGAG 3956
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1685 AspLeuGlyTyrgLylYsPheGlnSerAsnLeuGlnThrAsnAsnAl 1701
3957 CAATAATCCCGCGCGCGCTGCTGATTCAGGCAATTCAGGCAAGATCCGG 4006
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1701 alYsPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLysAlaP 1718
4007 CAGGTTTCGCGGATTCGCGCATCGAACCACATCGGCGCAACGCGGTAT 4056
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1718 heAsnLeuGlnAsnPheAlaValLysProThrValGlyValArgTySer 1734
4057 TTTCGTCAAAAAGCGATTACGATACGAAACGTCAATATCGCCGCC 4106
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```

1735 TyrLeuSerAsnAlaaspPheAlaLeuAlaGlnAspArgIleLysValAs 1751
4107 GGGCTTCATTCACACCGCTACCGGGCGCATTAAGGCAATATTCAT 4136
1751 nProIleSerValLysThrAlaPheAlaGlnValAspLeuSerLysThr 1768
4157 TCMAACCGCGGCACACATTCATTCACCTTATTTGAGCGCTCCAT 4206
1768 yrhSLenGlyGlu...PheSerIleThrProIleLeuSerAlaArgLys 1783
4207 ACCGATGCCGCTCCGCAAGTCGGAAGCGCGCTCAATACCGCGAT 4236
1784 ..AspAlaAsnGlnGlyAsnGlyLysIleAsnValSerValTyrAsp 1799
4257 GCGCGAGATTTGGCAAAACCGGAGTGGGAAATGGGCGGTAACGGCG 4306
1799 eAlaTyrAsnValGlnAsnGlnGlnIleTyrAsnAlaGlyLeuLys 1816
4307 AATTCMAAGTTTCACGCTCTCCACGCTGCCGCGCCGCAAGGGCGG 4356
1816 ySTyriHisnValLysLeuSerLeuIleGlyLysLeuThrLysAlaLys 1832
4357 CAATTGGACGGCAGACAGCGCGGCGCATCAATTAGCGCTAC 4398
1833 GlnAlaGlnLysGlnLysThrAlaGlnValLysLeuSerPhe 1846
seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-10661A-6

```

```

seq_documentation_block:
? Sequence 6, Application PCT/US9510661A
? GENERAL INFORMATION:
? APPLICANT: Washington University, et al.
? TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Flehm, Honbach, Test, Albritton & Herbert
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: United States
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US95/10661A
? FILING DATE: 16-AUG-1995
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/296,791
? FILING DATE: 25-AUG-1994
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Treacartin, Richard F.
? REGISTRATION NUMBER: 31,801
? REFERENCE/DOCKET NUMBER: FP-59941/RET
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 781-1989
? TELEFAX: (415) 398-3249
? TELEX: 910 277299
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1848 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
PCT-US95-10661A-6

```

alignment_scores: Quality: 1513.00 Length: 1916
 Ratio: 1.519 Gaps: 60

Percent Similarity: 51.983 Percent Identity: 25.418

alignment_block:
 US-09-303-518d-653 x PCT-US95-10661A-6 ..

Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

```

64 CGGTCGCGCGCGCTTACCATATATCGCTGCGGATTCGCG 113
   ::::::::::: ::::::::::: :::::::::::
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAAGCCGCGCGGACACACTTATTCGGCATCAATCAATATATC 163
   ::::::::::: ::::::::::: :::::::::::
21 oYrThrGlnAlaAlaLeuValArgAspAspValAspLysGlnIlePhe 38
164 GCGACTTTCGCGCAAAATAAGGCAAGTTCGACGCGCGCAAGATAT 213
   ::::::::::: ::::::::::: :::::::::::
38 rGaSPheAlaGlnLysAsnLysGlyLysPheSerValGlyValAlaThrAsnVal 54
214 GAGCTTTACACAAATAAGGAGGAGTGTGCGCAATTCATGACGAAC 263
   ::::::::::: ::::::::::: :::::::::::
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsn 71
264 C...CGATGATGATTTTCTGTGTATG...CGTAAGCGGTGGCG 307
   ::::::::::: ::::::::::: :::::::::::
71 yIleProMetIleAspPheSerValAlaAspValAspLysArgIleAla 88
308 CATTGGCGCGCGCATCAATATATTCGACGCGGCACAT...AAGCGCG 354
   ::::::::::: ::::::::::: :::::::::::
88 hrLeuValAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
355 TATAACAATGTCGATTTTGTGT...GCGAGGAGGAAGCA 389
   ::::::::::: :::::::::::
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAlaAsnGly 121
390 TCCCGATCAGCACCGC...TTTCTTACCAATG 421
   ::::::::::: :::::::::::
121 nAlaLysSerHisArgAspValSerSerGlnLysAsnArgTyrTyrThr 138
422 TGAAGAATAATATAT...AAGCA 444
   ::::::::::: :::::::::::
138 AlGlnLysAsnAsnPheProThrGlnAsnValThrSerPheThrLysGlu 154
445 GGGACTAAGCGGCATCTTATGGCGGATTCATATATCCGCGTTGCA 494
   ::::::::::: :::::::::::
155 GluGlnAspAlaGlnLysArgArgLysAspTyrTyrMetProArgLeu 171
495 CAATTGTTCACAGATGCAAGACCTGTGAG...ATGACCATGTATATG 541
   ::::::::::: :::::::::::
171 pLysPheValThrGlnValAlaProIleGlnLysAsnThrAlaAsnAsn 188
542 ATGGGTGGAATATCGCTGATTTAAATATACCTGATGCTTGGAAAT 591
   ::::::::::: :::::::::::
188 snLysGlnLysTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204
592 GGAGCAGGACAGACATAT...TG 611
   ::::::::::: :::::::::::
205 GlySerGlnThrGlnPheIleTyrLysLysGlySerArgTyrGlnLeu 221
612 GCGGCTGATGAAGACAGACCAATACCGCGAAAGTTATATCAT... 657
   ::::::::::: :::::::::::
221 eLeuThrGlnLysAspLysGlnLysAsnLeuArgAsnThrPaspValG 238
658 .....ATTGACGCGCATATTCGCTGCGTGGCT 687
   ::::::::::: :::::::::::
238 LysLysAspAsnLeuGlnLeuValGlyAsnAlaTyrThrLysGlnIleAla 254
688 GGCATATCTTTGCACAAATGATCAGTGGTGGCAGCATCACTATGAG 737
   ::::::::::: :::::::::::
255 GlyThrProTyrLysValAsnHisGlnAsnGlyLeuIleGlyPheGln 271
738 TAGCGAAAAAATTAACATAGC...CCATATGTTTATACACAGGA. 783
   ::::::::::: :::::::::::

```

271 yAsnSerLysGluGlnHisSerAspProLysGlyIleLeuSerGlnAsp 288
784GGCTATTGGCGACAGTGGCTCACCAGTATTATC 819
288 roleThrAsnTyrAlaValLeuGlnAspSerGlySerProLeuPheVal 304
820 TATGATGCCCAAAAGCAAAAGTGTATTAAATGAGTATTGCAACAGG 869
305 TyrAspArgGlnLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321
870 CAACCCCTATTAGAAAAACCATGGCTTCACCTAGTCTGTAAGATT 919
321 PALGlyTyr.....AsnLysLysSer 329
920 GGTTCATGATGAATCTTTCGTGAGATACCATTCAGTATTCTAGCAA 969
329 rPgInGlnTrpAsnIleTyrLysHisGlnPheAlaGlnLysIleTyrGln 345
970 CCACATCAAAATGGAAATACTTTTAAACGACAATAAATGCGCAGG 1019
346 GlnTyr.....SerAlaGln 350
1020 AAAATGCGCAACATAACACTATTCCTACCTTATAGATTAAAA 1069
350 ySerLeuIleLysSerAsnThrGlnTyrThrPglAlaThrGlySer 367
1070 CACGACCGCT.....CAATGTATTATGTTCTTATTCGAG 1107
367 hrSerThrIleThrGlyGlyGlnProLeuSerValAspLeuThrAsp 383
1108 ACACAGAGAACCTGTTTATCATGCTGACAGTGGGTCAACACTTATCG 1157
384 GlyLysAspLysPro..... 388
1158 ACCGAGCTGATATGAGAAATATTCTTATTGACAAAGAAAG 1207
389AsnHisGlyLysSerIleThrLeu.....LysGlySerG 400
1208 GTGATTCATGATTACGACCAACATCAACAGGCGGCGGTTCAT 1257
400 LThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyLeuPhe 416
1258 TTTGAGGTAATTTTACGGTC...TCGCTAAAAAACAAGAACTGGCA 1304
417 PheGlnLysAspTyrGlnValLysGlyThrSerAspSerThrTrpLys 433
1305 AGCGGCGGCGTTCATATCAGTGTGCGAGTACCGTTCATTGGAAAGTAA 1354
433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTrpLysValH 450
1355 ACGGCGTGGCAACGACCGCTGCCAAATCGCAAGGACGCGTCTG 1404
450 IsAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
1405 GTTCAGCCCAAGGGAACCAAGGCTCGTCAAGCTGGCGCAGCTTAA 1454
467 ValGlnLysGlyLysAsnGlnGlyLeuLeuLysValGlyAspGlyTh 483
1455 AGTCATTTGATGACGAGCGGAGCATCAAGCAAAAAACAAGCTTTA 1504
483 rValIleLeuLysGlnLysAlaAspAlaAsnLysValGlnAlaPheS 500
1505 GTCAAAATCGGCTTGTCAGCGGAGGAGGCAACTGAATGGCGAT 1554
500 ergInValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516
1555 AATCATGTTACCCGACAAACTTATTTGGCTTCGCGCGCAGCTTT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyLysArgLe 533
1605 GGAATTGAAGGGGATCGCTTCGTTCCACCCGATTCAAAATACCGATG 1654
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
1655 AAGGGGATGATTGTCACACCAATCAAGACAAGAAAGATCCACCTTACC 1704
550 spGlyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGCGCAATAAGATATTACTACA.....ACCGCAATTA 1742
567 IleThrGlyLysSerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1743 C..... 1743
583 nIleGlnAlaGlnAspAspAspHisProLeuArgIleArgSerIleProT 600
1743 1743
600 yTrpGlnLeuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
1743 1743
617 LysGlyAlaSerThrArgSerGlnLeuProGlnAsnSerGlyLysSerAs 633
1743 1743
633 nGlnAsnTrpLeuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1744AACACTTGATAGCAAAAAAGAAATGGCTACACAGGTGTT 1788
650 alMetAsnHisIleAsnAsnGlnArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGGAGAAAGATGCAACAAACG...AACGGGCGGCTCAATCTGAATTA 1835
667 GlyGlnGlnGlnThrLysAlaThrGlnAsnGlyLysLeuAsnValThrPh 683
1836 CCACCGGAAAGACCGGATCGCACTTACTGTTCCGGCGCAACAAT 1885
683 eaSnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyThrAsnL 700
1886 TAAACGCAATATCAGCAACAAACGCAAACTGTTTCCGCGCGCA 1935
700 euAsnGlnLysPheAsnValGlnLysGlyThrLeuLeuSerGlyArg 716
1936 CCGACACCGCACCGCTACATCATTTAGAGCGGGGTGTCAAAATGGA 1985
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 ArgT.....ATCCCAAGAGAAATCGTGGGACAAAGTTCGATCG 2029
733 pRoHisPheThrGlnAsnAsnGlnValAlaValGlnAspAspTyrIleA 750
2030 ACCGCACTTTAAAGCGGAAACCTCCATATTCAGGCGGACAAACCGGTG 2079
750 snArgAsnPheLysAlaThrThrMetAsnValThrGlyAsnLaserLeu 766
2080 GTTTC...CGCAATGTTCCAAAGTGAAGGCGATTCGATTTAAGCA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCGCAAGAGTTTGGTGGTGGCACCGCATGAAGCCACACAACT 2176
783 naSnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrValC 798
2177 GTACAGCTCGGAGCTGAGGAGCTGACAACTGTGACGAAATTAACCA 2226
798 yValAlaArgSerAspTyrThrGlyTyrValThrCysHisAsnSerLysLeu 814
2227 ACCGAGATAAAGTATGCTCATTTGAGCAAGACGACATCAGAGCA 2276
815 Ser...GlnLysAlaLeuAsnSerPheAsnProThrLysLeuAlaGlyAs 830
2277 TGTCAAGCTTCGGGATCAGCGTCAATTAATTCACAGAGACTGGCACAC 2326
830 nValAsnLeuThrGlnAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846

```
2327 TCACGGCAATCTTAGTCAGCGCGA..... 2352
||| |||
846 eupheglyThrIleGlnSerIleGlyThrSerGlnValAsnLeuGlu 862
2353 GACACGCATATACGTTACGGCGCAACGCCAC..... 2385
||||| |||
863 AsnSerHisThrPheThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
2386 .CAAAAGCGCACTCAGCCCTCGGGGCAACCCCA..... 2421
||||| |||
879 uThrAsnGlyHisIleHisLeu..... AsnAlaGlnAsnAspAlaAsnL 894
2422 .GCAACATTTATTCACACCATTAACGGCAACACATCGGCTTCG 2466
||||| |||
894 yValThrThrTyrAsnThrLeuThrValAsn..... SerLeuSer 907
2467 GACAAATGCTCATTT.....AATCTAAGCAA 2492
||||| |||
908 GlyAsnGlySerPheTyrTyrTrpValAspPheThrAsnAsnLysSerAs 924
2493 CAACGGCGTACAAACGCGCAGT..... 2514
||||| |||
924 nLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValA 941
2515 .....CTGACGCTTCCGACAAAC 2532
||||| |||
941 laAspLysThrGlyGluProAsnHisAsnGlnLeuThrLeuPheAspAla 957
2533 GCTAAGCA.....AACGTAAAGCATTCGCCACTCAACGGCAATGT 2573
||||| |||
958 SerAsnAlaThrArgAsnAsnLeuGlnValThrLeuAlaAsnGlySerVa 974
2574 CTCCCTACCGCATTAAGCAGCATTCATTTGAAACAGCGGCTTACG 2623
||||| |||
974 L.....AspArgGlyAlaTrpLysTyrLysLeuArgAsnValAsnG 988
2624 GAAAAATCAGCGCGCGCAAGATACGGCATTAACCTTAAAGACAGCAA 2673
||||| |||
988 LysArgTyrAsp.....LeuTyrAsnProGln 996
2674 TGAACGCTGCGGCGGCAAGCAATTGACATTTAAACCTTGACAAAGC 2723
||||| |||
997 .....ValGlnLysArgAsnGlnThrValAspProThr 1007
2724 CACCATTAACCTCAATTCGCCCTATCGACACAGATGCGGCGCGCAAA 2773
||||| |||
1007 rAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGlns 1024
2774 CCGGCGAGT..... 2781
||||| |||
1024 eTrAsnAsnGlnGluIleAlaIArgValGluThrProValProProAla 1040
2782 .....GCGGCAAGATGCGGCGCGCGCT..... 2805
||||| |||
1041 ProAlaThrArgLysSerAlaIleAlaSerGlnProGluThrAlaArgProAl 1057
2805 ..... 2805
1057 aGluThrAlaGlnProAlaMetGluGluThrAsnThrAlaAsnSerThrG 1074
2806 .....TCGCGC 2811
||||| |||
1074 LuThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSer 1090
2812 CGTTCCTATTCATTCGTTACG..... 2832
||||| |||
1091 GluSerValProSerGluThrThrGlyLysValAlaGluAsnProProG 1107
2832 ..... 2832
1107 nGluAsnGluThrValAlaLysAsnGlnGlnAlaThrGluProThr 1124
2833 .....CGGCAACTTGGCGAGAA 2850
||||| |||
1124 roGlnAsnGlyGluValAlaLysGluAspGlnProThrValGluAlaAsn 1140
2851 TCCCGTTTACACAGCGTGCAGGTAAC...GGCAATTTGAACGGTCAGG 2897
||||| |||
1141 ThrGlnThrAsnGluAlaThrGlnSerGlnGlyLysThrGlnGluThrG 1157
2898 AACATTCGCGCTTATGTCGAA..... 2919
||||| |||
1157 nThrAlaGluThrLysSerGluProThrGluSerValThrValSerGluA 1174
2920 .....CTCTTCGCGC 2928
||||| |||
1174 snGlnProGluLysThrValSerGlnSerThrGluAspLysValVal 1190
2929 TACCGGCGGCAAAATTCGAGCTG..... 2952
||||| |||
1191 GluLysGluGluLysAlaLysValGluThrGlnGluThrGlnLysAlaPr 1207
2952 ..... 2952
1207 oGlnValThrSerLysGluProProLysGlnAlaGluProAlaProGluG 1224
2953 .....GCGGAAGTTCGCAAGGCACT 2973
||||| |||
1224 LuValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGlnThr 1240
2974 TACACCTTGCGTGC..... 2988
||||| |||
1241 GlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLysPr 1257
2988 ..... 2988
1257 oAlaGluGluThrGlnGlnProSerGluLysThrAsnAlaGluProValIT 1274
2989 .....AACATTAACCGGCAACGCGTAAGTCTCAGCAA 3024
||||| |||
1274 hProValValSerGluAsnThrAlaThrGlnProThrGlnThrGlnGlu 1290
3025 TTGACGATGGAAGAAAGCAACACACCGCTGCGGAATCTTAA 3074
||||| |||
1291 ThrAlaLysValGluLysGluLysThrGlnGluValProGlnVal.... 1305
3075 TTTACACCTTGCAAAACGACAGCTGATGCGGCGGCGCTTATCAGC 3124
||||| |||
1306 .....AlaSerGlnGluSerProLysGlnGluGlnProAlaAl 1318
3125 TTATCGCAAAAGCGGCGAGTTCCGCTGCAATAACGGTCAAGAACAA 3174
||||| |||
1318 aLysProGlnAlaGlnThrLysProGlnAlaGluProAlaArgGluAsn 1334
3175 GAGCTTTCGACAAA..... 3189
||||| |||
1335 ValLeuThrThrLysAsnValAlaGlyGluProGlnProGlnAlaGlnProG 1351
3190 .....CTCGGCAAGCGGAGAACAGAGGCC...GCCT 3220
||||| |||
1351 nThrGlnSerThrAlaValProThrThrGlyGluThrAlaAlaAsnSerL 1368
3221 TGACGGCAAAACGACCAACTTGCAGCAACAAACAGCGCGAA..... 3264
||||| |||
1368 ySProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAla 1384
3265 .....AAAGCAACGCGCAAGGCTTGA 3287
||||| |||
1385 ArgGluAsnValSerThrValAsnThrLysGluProGlnSerGlnThrSe 1401
3288 CGGCGTGAATTCGCGCGGCGGCAATGCCAGCAAAAGAGAGTGTG 3337
||||| |||
1401 rAlaThrValSerThrGlnGlnProAlaLysGluThrSerSerAsnValG 1418
3338 CCGAACCGGCG..... 3348
||||| |||
```

1418 IuGIInProAlaProGIuAsnSerIleasnThrGIYSerAlaThrThMet 1434
3348 3348
1435 ThrGIInThrAlaGIuLySSerAspLysProGIInMetGIuThrValThrGI 1451
3349CGGAG.....GCAGCGGGGAAATCCG 3373
1451 uAsnAspArgGIInProGIuAlaAsnThrValAlaAspAsnSerValAlaA 1468
3374 GCATTAGCGAGCGGAGAGAGAAACCG.....GTGAG 3411
3412 GCGGATAAGACACCGCTTGGCAACAGCGGAGCGGAAACCGCGCC 3461
1485 GIInProGIuThrSerAlaGIuGIInThrValAlaSerThrGIInGI 1501
3462 GCGTACC.....ACGCGCTTCCCGCGCGCGCGCG 3493
1501 uThrThrValAspAsnSerValSerThrProLySProArgSerArgTyr 1518
3494 CCGCGCGGAGATTGCCGACCGCGACCGCCGCAACCGCAACCGCGAG 3543
1518 hrArgArgSerValGIInThrAsnSerTyrGIuProValGIuLeuProThr 1534
3543 3543
1535 GIuAsnAlaGIuAsnAlaGIuAsnValGIInSerGIuLyAsnAsnValAlaAs 1551
3544CGGAGCTGATCAGCGCTTATGCCAATAGCGGTT 3577
1551 nSerGIInProAlaLeuArgAsnLeuThrSerLyAsnThrAsnAlaValI 1568
3578 TGAGTGAATTTCCGCC.....ACGTCACAGCGCTTTC 3612
1568 IeSerAsnAlaMetAlaLysAlaGIInPheValAlaLeuAsnValGIuLyS 1584
3613 GCGGTACAGACGAAATGACCGCGTG...TTGGCGAGACCGCGCGCA 3659
1585 AlAlaLysGIInHisIleSerGIInLeuGIInMetAsnGIuGIInTyr 1601
3660 CCGCGTTTGCACAGCGGATCGGAGACCAACACTACCGTTTCGAG 3709
1601 rAsnValTyrIleSerAsnThrSerMetAsnLyAsnTyrSerSerGIuG 1618
3710 ATTCCGCGGCTACCGCCCAACAAACGACCTGGCCCAATCGGATTCAG 3759
1618 InTyrArgArgPheSerSerLySerThrGIInThrGIInLeuGIITyrPSP 1634
3760 AAAAACCTCGGACGCGG...CGGTCGCGATCCTGTTTCGACACACG 3806
1635 GIInThrIleSerAsnAsnValGIInLeuGIIGlyValPheThrTyrValAr 1651
3807 GACCGGAAACACTTCGACAGCGGATCGCAACTCGGACGCGCTGCCC 3856
1651 gAsnSerAsnAsnPheAspLysAlaSerSerLySAsnThrLeuAlaGIInV 1668
3857 ACGGTGCGGCTTTCGCGCATACGCGCATCGGAGTTCCAGATCGGATC 3906
1668 aLAsnPheTyrSerLySTyrTyrAlaAspAsnHisTyrTyrLeuGIITle 1684
3907 AGCGCGGCGCGGCTTTCAGCGGACGCGCTTCAGACGCGCATTCAGAG 3956
1685 AspLeuGIITyrGIuLySPhGIInSerAsnLeuGIInThrAsnAsnAl 1701
3957 CAATAACCGCGCGCGCTGCTGCATTACGCATTACAGGCAAGTACCGCG 4006
1701 aLySPhAlaArgHisThrAlaGIInLeuGIIThrAlaGIILysAlaP 1718
4007 CAGGTTCGCGGATTCGCGCATCGCAACGCGCATCGGCGCGCGCTAT 4056
1718 hAsnLeuGIInPheAlaValLySProThrValGIuAlaTyrTyrSer 1734

4057 TTCGTCAAAAAGCGGATTACGATACGAAGAAACGTCATATCGCCACCC 4106
1735 TyrLeuSerAsnAlaAspPheAlaLeuAlaGIInAspArgIleLySValS 1751
4107 GGGCCTTGCAATTCACCGCTACCGCGCGGCGCATTTAAGCAGATTATTCAT 4156
1751 nProIleSerValLySThrAlaPheAlaGIInValAspLeuSerTyrThrT 1768
4157 TCAAAACCGCGCACACATTTCCATTCAGCGCTTATTGAGCCTGTCTAT 4206
1768 yrhIsLeuGIInLy...PheSerIleThrProIleLeuSerAlaArgTyr 1783
4207 ACCGATCCGCTTCGCGCAAGTCGCAACGCGCGGTCATATACGCGTAT 4256
1784 ...AspAlaAsnGIInLySngLySIIeAsnValSerValTyrAspPh 1799
4257 GGGCAGGATTTTCGCAAAACCGCGAGTCGGAATGGGCGTAAACGCGCG 4306
1799 eAlaTyrAsnValGIuAsnGIInGIInLySThrAsnAlaGIILeuLySLeu 1816
4307 AATCAAGCTTTCACGCTGCTCCCTCCACGCTGCGCGCGCAAGGCGCG 4356
1816 ySTyrHisAsnValLySLeuSerLeuIleGIuGIIThrLySAlaLyS 1832
4357 CAATTGGAAGCGGACGACGCGGCGCATCAATTAAGCTAC 4398
1833 GIInAlaGIuLySngLySThrAlaGIuValLySLeuSerPhe 1846
seq_name: /cgn2_6/plodata/1/lae/5B.COMB.pep:US-08-728-470-9
seq_documentation_block:
Sequence 9, Application US/08728470
Patent No. 5928651
GENERAL INFORMATION:
APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,470
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 9:


```

1334 GTACCGTTACTGGAAAGTAAACGGC..... 1359
      ::::::::::::::::::::
489  lyservalasnphelysleuasnalasersersasnileglnthrpro 505
      ::::::::::::::::::::
1360 .....GTGGCAAGACCGCCTGTCCAAATCGCCAAAGCAGCCT 1400
      ::::::::::::::::::::
506 glvalailelleysserglasnphasnvalserglyserthrle 522
      ::::::::::::::::::::
1401 GCTGCTCAGCCCAAGG.....GAAACC 1426
      ::::::::::::::::::::
522 uasnleuysalaglulyserrhrthrthlapheserilleglnsna 539
      ::::::::::::::::::::
1427 AAGCTCGGTGACGGTGGCGACGTAAGTCATCTAGATCAG...CAG 1473
      ::::::::::::::::::::
539 spleuasnleuasnalarhrthrlyglasnilethrilearglnvalglu 555
      ::::::::::::::::::::
1474 GCGGACGATCAA..... 1485
      ::::::::::::::::::::
556 glythrpserserhrgvalasnlysglyvalalaalyslysasnileth 572
      ::::::::::::::::::::
1486 .....GGCAAAAACAAGCCTTTAGTGAAA 1510
      ::::::::::::::::::::
572 rphelysglylyasnilethrphleglysergllysalathrthrglut 589
      ::::::::::::::::::::
1511 TCGGCTTGTCACGGCGACGGGACGGTGCATGTAATGCCGATATCAG 1560
      ::::::::::::::::::::
589 le.....Lysglyasnvalthrileasnlyasnthrpsn 600
      ::::::::::::::::::::
1561 TTCAACCCCGCAAACTCTATTTCGGCTTCGGCGGACGCTTGGAATT 1610
      ::::::::::::::::::::
601 Alathr.....Leuarglyalasnphelaagl 610
      ::::::::::::::::::::
1611 GAACGGGCGATTGCTTCGTTCCACCGCATTCAAAATACGATGAAGGG 1660
      ::::::::::::::::::::
610 uasnlyserproleuasnile.....Alaglya 620
      ::::::::::::::::::::
1661 CGAGATTGTCAACCAAT...CAAGCAAGATCCACCGTTACCATTT 1707
      ::::::::::::::::::::
620 snvalileasnnglyasnleuthrthrthlaglserilleasnile 636
      ::::::::::::::::::::
1708 ACAGGCATATAAGATTACTACCAACCGCAAT..... 1740
      ::::::::::::::::::::
637 Alaglyasnleuthrvalserlygllyalasnleuclnalailethrps 653
      ::::::::::::::::::::
1741 .....AACCAACTTGATAGCAAAA 1762
      ::::::::::::::::::::
653 nlythrphasnvalalaglyserphasnpsasnnglyalaserasn 670
      ::::::::::::::::::::
1763 AAGAATTGGCTACACGGTTGGTTGGCGAAGATGCAACCAACG 1812
      ::::::::::::::::::::
670 leserilleala...Argglylyalalysphelysaspilleasnthr 685
      ::::::::::::::::::::
1813 AACGGGGGCTCAATCTGAATTACCAACCGAAGAACGGATCGCACTTT 1862
      ::::::::::::::::::::
686 serser...Leuasnilethrthrasnseraspthrlyrthrlythr 701
      ::::::::::::::::::::
1863 ACTGCTTCGGCGGACAAATTAAACGCAATATACGCAACAACAAG 1912
      ::::::::::::::::::::
701 e.....ileysglyasnileserasnlyserg 711
      ::::::::::::::::::::
1913 GCAAACTGTTTTTCAGCGGACAGACCGGACGCTTACATCATTTA 1962
      ::::::::::::::::::::
711 lyspleuasnilelleasplyslyser...Aspalaglulieglille 726
      ::::::::::::::::::::
1963 GGAAGCGGGTGTCAAAATGAAGGTATCCACAAAGAGAAATCGTGTG 2012
      ::::::::::::::::::::
727 glylyasnileserclynlyglu.....glyasnleuthr 739
      ::::::::::::::::::::
2013 GGACAGCATTTGATGAC.....CGCATTTTAAAGCGGAAA 2050
      ::::::::::::::::::::
739 eserserapslyvalasnilethrasnnglnilethrilelysala.... 754
      ::::::::::::::::::::
2051 ACTTCATATTTCAGGGCGACAAGCG..... 2076
      ::::::::::::::::::::

```

```

755 .....glyvalileuglyglyargserpserserthrthlaleuasnala 769
      ::::::::::::::::::::
2077 .....GTGGTTCCCGCAATGTTGCCAAAGTGAAGCGGATTTGCATT 2120
      ::::::::::::::::::::
770 Asnleuthrilleglnthrlysglyleuylsleuallaglyasnleuasn 786
      ::::::::::::::::::::
2121 AAGCAATCACGCCCAAGCACTTTTCGCTGCGACCGCATCAAGCCACA 2170
      ::::::::::::::::::::
786 eserglyphasnlysalagluliethrthralatlysasnnglyserasp 803
      ::::::::::::::::::::
2171 CAATCTGACACGCTTCGAGCTGACGCGTCTGACAGTTGTACCGAAAA 2220
      ::::::::::::::::::::
803 hrillegllyasnalasergly.....Glysnalalaspalalys 815
      ::::::::::::::::::::
2221 ACCATTACCGACGATAAGTATGATGCTTCATTGACGACGACATPCAG 2270
      ::::::::::::::::::::
816 Lysvalthrphasnplysvallyasnpserslylleserthrps...gl 831
      ::::::::::::::::::::
2271 AGGCATGTCAGCCTTCGCGATCAGCCTCATTTAAATTCACAGACTTG 2320
      ::::::::::::::::::::
831 yhlasnvalthrleuasnserglu.....ValL 841
      ::::::::::::::::::::
2321 CCACACTCAACGCGCAATCTTAGTGACGCGGAGAC.....ACGCATAT 2364
      ::::::::::::::::::::
841 ythrserasnnglyseraserasnalaglyasnpsasnserthrlyleu 857
      ::::::::::::::::::::
2365 ACGGTTACG...CGACACGCCACCAAAACGGCACTCAGCCTCGTGSG 2411
      ::::::::::::::::::::
858 Thrileserlaleysaspvalthrvalasnasnvalthr..... 871
      ::::::::::::::::::::
2412 CAATGCCCAAGCAATTAATCAAGCACATTAACGGCAACACATCGG 2461
      ::::::::::::::::::::
872 ...SerhlslysthrileasnileseralalalaglyasnvalthrT 887
      ::::::::::::::::::::
2462 CTTGGACAAATGCTTCATTTAATCAACCAACGCCGTA..... 2502
      ::::::::::::::::::::
887 hrlysglythrthrthrileasnlatthrthglyseralgluvalthr 903
      ::::::::::::::::::::
2503 .....CAAAACGCCAGTCT 2516
      ::::::::::::::::::::
904 Alaglnasnnglythrilleysglyasnilethrserglasnvalthrps 920
      ::::::::::::::::::::
2517 GACGCTTTCGACCAAC.....GCTPAGGCAACGTAAGCATTCGCCGC 2560
      ::::::::::::::::::::
920 lthrthlathrgluasnleuvalthrthrlyasnalaileasnlat 937
      ::::::::::::::::::::
2561 TCACGGCAATGTCTCCTAGCCGATAGGCGAGTATTCATTTTGAAAAC 2610
      ::::::::::::::::::::
937 hrserglythrvalasnileserthrly..... 946
      ::::::::::::::::::::
2611 AGCGCTTACCGGAAATCAAGCGGCGGACAGATACGGCA..... 2652
      ::::::::::::::::::::
947 .....Thrlyaspilleysglylyllegluserthrserglyasn 960
      ::::::::::::::::::::
2653 ..TTACACTTAAAGACAGGAATGAGCGCTGCGGCGACAGGAATAG 2701
      ::::::::::::::::::::
960 nvalasnilethrthlaserglyasnthrleu.....LysvalS 973
      ::::::::::::::::::::
2702 GCAATTTAAACCTTGACAAACGCCACATTTACCTCAATTCGCGCTATGCA 2751
      ::::::::::::::::::::
973 erasnilethrlyglinaaspvalthrvalthralasp..... 985
      ::::::::::::::::::::
2752 CACGATGCGGACGCGCAACCGGACGAGTGGGAGATGCGCGCGCG 2801
      ::::::::::::::::::::
986 .....Alaglyalaleuthrthrthralagly..... 994
      ::::::::::::::::::::
2802 CGGTCGCGCGCTCCATTATTCGTTACGCGGCAACTTCGCGCAAT 2851
      ::::::::::::::::::::
995 .....serthrileseralathrthrlysnala..... 1004
      ::::::::::::::::::::
2852 CCGGTTTCAACAGCGTGAACGGTAATTAACGGTACAGGTACGAGACA 2901
      ::::::::::::::::::::

```

```

1005 .....AsnIleThrThrLysThrGlyAspIleasnGlyLys..... 1016
2902 TTCGGCTTTATGTCGAACTCTTCGGCTACCGCAGCGCAAAATTGAAGCT 2951
1016 ..... 1016
2952 GGGGGAAGATTTCGAAAGCATTAC.....ACCT 2980
1017 ValGluSerSerSerGlySerValThrIleuValAlaThrGlyAlaThrL 1033
2981 TGCGTGTCAACATACCGCAGCAAGAACCCCTAGTCTCCGACGCAATTGACG 3030
1033 euAlaValAlGlyAsnIleSerGlyAsnThrValThr.....LleThr 1046
3031 GTAGTGAAGAAAAGACACACACCGCTCCGAAAATCTTAATTTCAC 3080
1047 AlaAspSerGlyLysLeuThrSerThrValGlySerThrIleasnGlyTh 1063
3081 CTTGCAAAAGAACACGTCGATGCCGCGCATGCGCTTATCAGCTTATCC 3130
1063 T..... 1063
3131 GCAAAAGCGGAGATTCCGCTGCATATATCCGTCGCAAGACAGAGCTT 3180
1064 .....AsnSerValThrThrSerSerGlnSer 1072
3181 TCCGCAAAACTCGGCAAGGGCGGGAAGAACAGAGCGCTTGAAGCGCAA 3230
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAla 1089
3231 ACAGCGCACACTT.....GCCGCAAAACACAGCGCGGAAAAG 3268
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105
3269 ACAACGGCGCAAAAGCCTTGACGCGCTGATTCGGCGCGCGCAATGCCACC 3318
1105 snGlyAlaAlaThrLeuThrAla....GluSerGlyLysLeuThrThr 1119
3319 GAAAGGCGAGAAAGTGT.....GC 3338
1120 GlnThrGlySerSerIleThrSerSerAsnGlyLthrThrLeuThrAl 1136
3339 CGAACCGCGCGGAGCGAGCGGGGAAAATGCC..... 3372
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153
3373 .....GCCATTATGACGCGCGAGAGAGAAAGAAACGGGTCCAGCGC 3414
1153 snThrThrGlyThrLeuThrThrGlyAspSerLys...IleAsnAla 1168
3415 GATTAAGACACCGCTTGCGAAGACGCGCAACCGGAAACCGCGCGGC 3464
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185
3465 TACGACCGCGCTTCCCGCGCGCGCGCGCGCGGAGATTTCGCGCAAC 3514
1185 aAlaSerGly..... 1188
3515 CGGAGCCCAACCGCAACCCCAACCGCAGCGCGCATCATCAGCCGTTAT 3564
1189 .....AspArgThrValAlaAsnAlaThr 1196
3565 GCCAATACGGGTTGAGTGAATTTCCGCCACGCTCAACACGCTTTCGC 3614
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValAs 1213
3615 CGTACAGACGAATTGAC.....CGCGTGTTCGCGAAG 3649
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleLeSerGluA 1230
3650 ACCGCGCGAAGCGCGTTGGACAGCGGCATCCGG...GACACCAACAC 3696
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysThr 1246

```

```

3697 TACGCTCGCAAGATTTCGCGCGCTACCGCCACACACCGCACTGCGCA 3746
1247 Ile.....GlnPro.GlyValAlaIleSerValGluGluValIleGluAlaL 1261
3747 AATCGTATGCGAGAAAAC.....TCG 3769
1261 ysArgValIleuGluLysAspLeuSerAspGluArgGluThr 1277
3770 GCAGCGGCGCGCTCGCATCCGTTTTCGACACACCGGACCGGAAACAC 3819
1278 LeuAlaLysLeuGlyValSerAlaValAlaArgPheValGluProAsnAl 1294
3820 TTCAGCAGCGCATCGCACTCGGACGCGCTTGCCCGACGCTGCCGTTT 3869
1294 alleThrVal.....AsnThrGlnAsnGluPheThrThrLysProSerS 1309
3870 CGGCAATACGCGCATCGCAGGTTCGACA.....TCGGCATACGCGCG 3913
1309 ergLlnValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325
3914 GCGCGGCTTTAGTAGCGGACGC 3937
1326 AlaArgValCysThrAsnValAla 1333

seq_name: /cgn2_6/prodata/1/iaa/6B_COMB.pep:US-08-719-641-9
seq_documentation_block:
; Sequence 9, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barksstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

```

;          TOPOLOGY: linear
US-08-719-641-9

```

```
alignment_scores:
```

Quality:	308.50	Length:	1542
Ratio:	0.439	Gaps:	75
Percent Similarity:	45.525	Percent Identity:	20.169

alignment_block

US-09-303-518D-653 x US-08-719-641-9 .

Align seg 1/1 to: US-08-719-641-9 from: 1 to: 1338

1	ATGAAAACACCCAAAGGCAACCCGAACACCCGAAGCCCTAA	50
67	MetIethrGlysprYsValThrIeuysThrGlyAlaValIleasPle	83
51	AACGGCGCCGATCCGCTTCTCGCCCGCTACTTACCAATATGCTGC	100
83	uSerGlyLys.....	86
101	TGCGCATTTGCCCCAGCCGGGCGACACTATTATGGCATCAAC	150
87GluGlyGlyGluThrTyrLeuGly.....	94
151	TACCAATACTATCGCATTTGGCGAAATAAAGCAAGTTTGCAGTCG	200
95GlyAspGluArgGlyu.....G1	104
201	GGCGAAAGATATGAGTTTACAAACAAAAGGCGATTTGGTGGCAAT	250
101	YLyAsnGlyIleGluIleuAlaLysLys.....ThrT	112
251	CGATGACGAAGACCCCGATGATGTGATTTTCTGTGTGTCGCGTAACGC	300
112	hTrIeuGluYsGlySerThrIleAsnValSerGlyLysGluLysGly	128
301	GTCGGCGCATTGGCGGCGCATATATATTGCG.....	333
129	ArgAlaIleValIleProGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl	145
334AGCGTGGACATATACGCGCGC.....	354
145	agInGlyLysAspIleAlaLysThrGlyLysIleValGluThrSerGlyu	162
355	..TATACAAATGTGTGATTTTGGTGGCGAGGACGATCC.....	393
162	IstYrIeuSerIleAspAspAlaAlaIleValLysThrLysGluTrpLeu	178
394	..GATAGACACCGCTTTCTTACCAAAATTGGAACAAATATTTTAA	410
179	LeuAspProGluAsnValThrIleGluAlaProSerAlaSerATGValG1	195
441	AGCAGGACTAACGGCGCATCTTATGGCGCGATATCATATGCCGGTT	480
195	uLeuGlyAlaAspArgAsnSerHisSerAlaGluValIleLysValThrL	212
191	TGCACAATTTGCADAAGATGCGAAGACCTGTTGAGATACCAATTATATG	540
212	eulysLysAsnAsnIleThrSerLeuThrIleuThrIleAsnThrIleSer	228
541	GATGGGAGCAAAATACGCTATTAATTAATATACCGTATCGTGTGCAT	590
229	AsnLeuLeuLysSerLalaHisValAlaLsnIleThrAlaAlaGlyLysLe	245
591	C.....	594
245	uThrValAsnSerIleSerIleGluArgGlySerHisLeuIleLeuH	262
592GGAGCAGACAAATATTTGGCGGTGATGAGAACAGAACCC	633
262	IstSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp.....	276

634	ATAAACCCGGAAGATTCAATATCATATTCATATTCAGACCCGATATTCGTGGCTGT	683
277IleThrSerGlu.....	280
684	CGGTGGCAATACCTTTTGCACAAATGATCAGGTGT	720
281GlyValAsnIleThrIleTyr.....SerGlyIleTyrValAspValH	295
721	..GGCAGAGTCACACTAGTAGTGGCGAAAAATTAAACATAGCCCATATGT	768
295	..ISLYASnIleThrLeuGlySer.....Gly	303
769	TTTTTACCAACAGAGAGCTCATTTGGCGAC	800
304	PheLeuAsnIleThrThrLysGluLysPheIleAlaPheGluAspLysSe	320
801	TGGCTCCACCAATGTTTACTATGATAGCCCAAAAGCAAAAGTCGTTAATA	850
320	..GlyIleArgAsnAsnLeuThrIleThrAlaGln.....	330
851	ATGGGATATTGCAACAGCAGCAACCCCTATTATAGAAAAAGCAATGCTTC	900
331	..GlyThrIleThrSerGlyAsn.....SerAsnGlyPhe	341
901	CAGCTAGTTCTGTAAGAATGTTGTTCTATGATGAAATCTTTGCTGGAGATAC	950
342	Arg.....PheAsnAsnValSerLeuAsnSerLe	351
951	CCATTCCAGATTCTACGAACCAACCATCAAAATGGGAAATACTTTATTACG	1000
351	u.....GlyGlyLysLeuSerPheMetA	359
1001	ACAATATATATGCGCGCAGGAAA	1023
359	spSerArgGluAspArgGlyArgThrLysGlyAsnIleSerAsnLys	375
1024ATCCATGCCAAMCA	1037
376	PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl	392
1038	TAAACACTATCTCTACCTTATAGATTAAACACGACCGTT	1080
392	..AspLysValSerThrPheTyrArgAspLysGlyArgThrTyrTrpAsnV	409
1081CAATGTTTAAATGTTCTTTATCC	1104
409	..AlaThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp	425
1105	GAGACGAGCAAGAGACCTGTTTATCATCTGTCAGGTGGGCAACAGTAA	1154
426	SerThr.....GlySerGlySerTh	432
1155	TCGACCCAGACTGAATAATGGAGAAATATTTCTTATATGACAAAGAA	1204
432	..GlyPheSerIleArgAsnAlaGlu..LeuAsnGlyIleThrPheAsnL	448
1205	AAGGTGAATTCATACTTACCCAGCAATCAACCAAGCGCGCGCGCTTG	1254
448	..GlyAla.....ThrPheAsnIleAlaGlnGlySer.....	457
1255	TATTTTGGAGGTAATTTTACGCTCGCCT	1289
458ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe	472
1290	CACGCAAACTGTGGCAA	1333
472	..TAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyG	489
1334	GAACCGTACTTGGAAACTAAACGC	1359
489	..LysValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnThrPro	505


```

1016 ..... 1016
2952 GCGCAAGATTCGAGCAGCACTTAC..... ACCT 2980
1017 ValGluSerSerSerSerValThrLeuValAlaThrGlyAlaThrL 1033
2981 TGCGTGTCAACATATACGGCAACAGACCCGTAAGTCTGAGCAATTGACG 3030
1033 euAlaValGlyAsnIleSerGlyAsnThrValThr..... IleThr 1046
3031 GTAGTGAAGAAAGAACACACACCCGCTGCGAAATCTTAATTTCAC 3080
1047 AlaSerSerGlyLeuThrSerThrValGlySerThrIleAsnGlyThr 1063
3081 CTGCAAAAGAACACGTCGATGCGCGCGATGCGTTATCAGTTATCC 3130
1063 T..... 1063
3131 GCAAAAGCGCGAGTTCGCTGCATATATCCGTCAAAGAACAGACTT 3180
1064 ..... AsnSerValThrThrSerSerGlnSer 1072
3181 TCGCAAAACTCGGCAAGGGGGAACAGAGCCCGCTTGACGGCAAA 3230
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAla 1089
3231 ACAGGCACAACTT..... GCGCGCAAAACAGCGGGAAG 3268
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla... LysA 1105
3269 ACAACGGCAAGCCTTGACGCGCTGATTCGCGCGCGCAATGCCAC 3318
1105 snGlyAlaIaThrLeuThrAla..... GluSerGlyLeuThrThr 1119
3319 GAAAGGCACAAAGTGT..... GC 3338
1120 GlnThrGlySerSerIleThrSerAsnGlyGlnThrThrLeuThrAl 1136
3339 CGAAGCGCGCGCGAGCGGCGGGAATGCC..... 3372
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaIaAsnValThrLeuA 1153
3373 ..... GGCATTATGACAGCGGAGAGAAAGAAAAACGGGTGACGGC 3414
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys... IleAsnAla 1168
3415 GATAAGACACCGCCTTGCGAACAACCGGAGCGGAAACCGCGCGCGC 3464
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185
3465 TACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCCAAC 3514
1185 aAlaSerGly..... 1188
3515 CGAGGCCCAACCGCAACCCCAACCGCAGCGGACTGATCAGCCGTAT 3564
1189 ..... AspArgThrValAlaValAsnAlaThr 1196
3565 GCCAATAGCGGTTTGAGTAATTTCCGCGACGCTCAACAGGTTTCCG 3614
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerValAs 1213
3615 CGTACAGAGCAATTGCAC..... GCGGTGTTTGGCGAAG 3649
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230
3650 ACCGCGCAAGCGCTTTGGACAAGGCGATCCGG... GACACCAACAC 3696
1230 smGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246
3697 TACCGTTCCCAAGATTCCGCGCTACCGCAACAAACGACCTGCGGCA 3746
1247 Ile..... GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1261

```

```

3747 ATCGGTATGCAAGAAACC.....TCG 3769
1261 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1277
3770 GCAGCGGCGCGCTCGCATCTGTTTTCGACACACCGGACGGAACACC 3819
1278 LeuAlaLysLeuGluLysAlaSerAlaValArgPheValGluProAsnAl 1294
3820 TTCGACAGCGCATCGCACTCGGCGACGCTTGCCACCGGTGCGGTTT 3869
1294 aIleThrVal..... AsnThrGlnAsnGluPheThrThrLysProSer 1309
3870 GCGGCATATCGCATCGCAGCTTGACAC..... TCGCATCAGCGCGG 3913
1309 eArgLysValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325
3914 GCGCGGTTTGTAGTCGCGCAGCC 3937
1326 AlaArgValCysThrAsnValAla 1333

```

seq_name: /cgm2_6/ptodata/1/iae/5B_COMB pep: US-08-617-697-9

seq_documentation_block:

Sequence 9, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Matzare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

STREET: Bldg. 1

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617,697

FILING DATE: 01-Apr-1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/302,832

FILING DATE: 05-Oct-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166

FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Berkstresser, Jerry M

REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER: 1038-557

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 1599 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-617-697-9

alignment_scores:

Quality: 308.50

Ratio: 0.439

Length: 1542

Gaps: 75

Percent Similarity: 45.525 Percent Identity: 20.169

alignment_block:
US-09-303-518d-653 x US-08-617-697-9 ..

Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599

```
1 ATGAAACACCGCACAAACGACAAACGACAAACACCGCAACCCCTAA 50
  ||| ||| ||||| ||| ||| |||
327 MetIleThrIleGlyAspIleValThrIleValThrGlyAlaValIleAspIle 343
51 AACCGCGCGCATCCGCTTCGCCCGCTTACTTACCATATGCGCTGCT 100
  ::|||:::
343 uSerGlyLys..... 346
101 TCGCATTCGCCCCAAGCCCGGGGAGACACTTATTTGCGCATCAAC 150
  ::|||::: ||||| ||||| |||||
347 .....GluIleGlyIleThrIleGly..... 354
151 TACCAATACTATCGCGACTTTGCGAAATTAAGCAAGTTTGCAGTCG 200
  ::|||::: ||||| ||||| |||||
355 .....GlyAspGluArgIleGly.....Gly 361
201 GCGCAAGATATTGAGTTTACAAACAAAAGGAGTGTGCGCAAT 250
  ::|||::: ||||| ||||| |||||
361 YLysAsnGlyIleGluIleValIleLys.....Thr 372
251 CGATGACGACAGCCCGCATGATGATTTCTGTGCTATGCGGTAACGC 300
  ::|||::: ||||| ||||| |||||
372 hIleuGluLysGlySerThrIleAsnValSerGlyLysGluLysGly 388
301 GTGGCGCATTTGGCGGCGATCAATATTTGTG..... 333
  ||| ::|||::: |||||
389 ArgAlaIleValIleThrGlyAspIleAlaIleValIleAspIleAsnAl 405
334 .....AGCGTGACATTAACGCGCG..... 354
  ::|||::: ||||| ||||| |||||
405 agIleGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGly 422
355 ..TATTAACATTTGATTTGTGTCGGGAGGACCAATCC..... 393
  ||| ::|||::: ||||| |||||
422 IStyIleSerIleAspAspAsnAlaIleValIleLysThrLysGluThrLeu 438
394 ..GATCAGCAGCCGCTTTCTTACCAATTTGCAAAAGAAATATTA 440
  ||| ::|||::: ||||| |||||
439 LeuAspProGluAsnValIleThrIleGluAlaProSerAlaSerArgValG 455
441 AGCAGGAGCTAACGCGCATCTTATGCGGCGGATATCATATGCCGCGT 490
  ::|||::: ||||| ||||| |||||
455 uLeuGlyAlaAspArgAsnSerIleSerAlaGluValIleLysValThrL 472
491 TGCACAAATTTGTACAGATGACAGAACCTTTGAGATGACGACGATAT 540
  ||| ||| ||||| ||||| |||||
472 euLysLysAsnAsnThrSerIleThrIleThrIleThrIleSer 488
541 GATGGGTGGAATAACGCTGATTAATAATAATATCCCTGATCGTGTCAAT 590
  ::|||::: ||||| ||||| |||||
489 AsnIleuLysSerAlaIleValIleValIleThrAlaArgAlaGlySer 505
591 C..... 591
505 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleuH 522
592 .....GGACAGAGCAGACATATTGCGGCTGTGATGAGACGAC 633
  ||| ||| ||||| ||||| |||||
522 IAspGluGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIle 536
634 AATAACCGCAAGAGTTCATATTCATATTCGCAAGCGCATATTCCTG 683
  ||||| |||||
537 .....IleThrSerGlu..... 540
684 CGGTGGCATATCTTTGCAAAATGATCGAGTGGT..... 720
  ||||| ||||| ||||| |||||
```

```
541 .GlyLysAsnLeuThrIleTyr.....SerGlyGlyThrPValAspValH 555
721 ..GGCAGAGTCATCTAGTAGTCGCAAAATTAACATATGACCCATATGCT 768
  ::|||::: ||||| ||||| |||||
555 IAspAsnIleThrIleuGlySer.....Gly 563
769 TTTTACCAACAGAGAGCTCATTTGGCGAC.....AG 800
  ||||| ::|||::: ||||| |||||
564 PheLeuAsnIleThrThrLysGluLysPheIleAlaPheGluAspLysSe 580
801 TGGCTCACCAGATGTTATCTATGATGCCCCAAAGCAAAAGGTTAATA 850
  ||||| ::|||::: ||||| |||||
580 rGlyArgAsnAsnLeuThrIleThrAlaGlu..... 590
851 ATGGGTATTCGCAACAGCAGCAACCCCTATATAGAAAAAGCAATGCTTC 900
  ||||| ::|||::: ||||| |||||
591 ..GlyThrIleThrSerGlyAsn.....SerAsnGlyPhe 601
901 CAGCTAGTTCTGAAGATTGTTCTATGATGCAATCTTGTGCGAGATAC 950
  ::|||::: ||||| ||||| |||||
602 Arg.....PheAsnAsnValSerLeuAsnSerLe 611
951 CCATTCAGTATCTACGAAACCATCAAAATGCGAAATATCTTTTAACG 1000
  ::|||::: ||||| ||||| |||||
611 u.....GlyGlyLysLeuSerPheIleThr 619
1001 ACAATATATATGCGCGCAGCAAAA..... 1023
  ||||| ::|||::: ||||| |||||
619 sPserThrGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 635
1024 .....ATCGATGCCAACA 1037
  ||||| ::|||::: ||||| |||||
636 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 652
1038 TAAACACTATCTCTACCTTATGATTAACAGCAACCGTT..... 1080
  ||||| ||||| ||||| |||||
652 aProLysValSerThrPheThrArgAspLysGlyArgThrIlePheAsnV 669
1081 .....CAATTGTTAATGTTCTTTATTC 1104
  ||||| ::|||::: ||||| |||||
669 aIleThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp 685
1105 GAGACGCAAGAGAACCTTTATCATGCTGCAGGTGGGCTCAACAGT 1154
  ::|||::: ||||| ||||| |||||
686 SerThr.....GlySerGlySerThr 692
1155 TCGACCCAGATGATATGAGAAAATTTCTTTATTCGCAAAAGCA 1204
  ||||| ::|||::: ||||| |||||
692 rGlyProSerIleArgAsnAlaGlu..LeuAsnGlyIleThrPheAsnL 708
1205 AAGGTGAATGATCTTACCAGCAACATCAACCAAGCGCGCGGCTTGG 1254
  ||||| ::|||::: ||||| |||||
708 ySAla.....ThrPheAsnIleAlaGlySer..... 717
1255 TATTTGAGGATATTTTACGGCTTCGCT.....AAAA 1289
  ::|||::: ||||| ||||| |||||
718 .....ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe 732
1290 CAACGAAACGTGCAAA.....GGCGCGGCGTTCATATACGTGATGCA 1333
  ||||| ::|||::: ||||| |||||
732 rAsnAlaAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyGlyG 749
1334 GTACCGTTACTTGAAAGTAACGCG..... 1359
  ::|||::: ||||| ||||| |||||
749 ySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGluThrPro 765
1360 .....GTGCAACAGCAGCGCGCTGTCCAAATGCGCAAGGACGCT 1400
  ::|||::: ||||| ||||| |||||
766 GlyValIleLeuLysSerGluAsnPheAsnValSerGlyLysThrIle 782
1401 GCTGTTCAAGCAAGGG.....GAAAC 1426
  ||||| ::|||::: ||||| |||||
782 uAsnLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGluAsnA 799
```


178 AATAAGGCAAGTTGCGAGTCGCGGCAAGATATATGAGTTTACACAA 227
199 AsnHsGlyLeuIleHnValGly...LysAspGlySerVal... 211
228 AAAAGGAGATTGGTCGCAAAATCGATGCAAGGCCCGATGATGATT 277
212AsnLeuIleGlyLysValLysAsnGlu..... 221
278 TTTCTGCGTATCCGCTAAGCC.....GTGGCGCATTTGGCGGCAT 321
222GlyValIleSerValAsnGlySerIleSerLeuLeuAlaGlyGln 237
332 CAA.....TATATGTGAG 335
238 LysIleHnIleSerAspIleIleAsnProHnIleHnTyrSerIleAla 254
336 CGTGGCAATACAGCGCGCTATACAAATGTTGATTTGGTCGAGGAA 385
234 aaIaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLysGly 271
386 GCATCCCGATCAGACCGCTTTCTTACCAATGTGAAAAAATATAT 435
271 LysAsn..... 272
436 TATAAGCAGGAGACTAAGCCATCGTTATGGCGGATATCATATGCC 485
272 272
486 GCGTTGCAAAATTTGTCAACATGCAAGACTGTGAGATGACCAGTT 535
272 272
536 ATATGATGGGTGAATAAGCTGATTTAATTAATACCTGATCGTT 585
273IleAsn.....Val 275
586 CGATGCGAGCAGCAGACAAATATTTGGCGGTCTGATGAAGACGACCA 635
276 ArgAlaIleHnIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 292
636 TAAACCGGAAGTTCAATCATATTCAGACCGCATATCTTGG.....C 679
292 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGluAlaG 309
680 TCGTCGCGCATFACCTTTGCACAAAATGATCAGT...GGTGGACA 726
309 IuIleGlyGlyValIleSerAlaGlnAsnGlnAlaLysGlyLys 325
727 GTCAACTTAGTAGCGAAATAATTAACAATAGCCCATATGTTTTTACC 776
326 LeuMetIleHnGlyAspLysValIleHn.....LeuLys 336
777 AACAGAGGCTCATTTGGCGACAGGCTCACCAATGTTATCTATGATG 826
336 sThGlyAlaValIleAspLeuSerGlyLys..... 346
827 CCAAAAGCAAAAGTGTATTAATGAGGGTATTCGAACAGCAGCAACC 876
347GluGlyGlyLysThr 351
877 TATATAGA.....AAAGCAATGCTTCCACCTAGT 908
352 TyrLeuGlyLysAspIleArgGlyGluGlyLysAsnGlyIleGlnLeuAl 368
909 TCGTAAAGATTGGTCTATGATGAATCTTTGCTGAGATACCATTCAG 958
368 aLysLys.....ThrThrLeuGluLysGlySerHnIleAsnV 381
959 TATTTACGAACCATCAAAATGAGAA..... 987
381 al.....SerGlyLysGluLysGlyLysArgAlaIle 391
388TACTTTTTTAAGCAATATATATATGCGCAGGAA 1022

392 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySe 408
1023 AATGATGCCCAATTAACCATATCTCTACCTTATACATTAACAAAC 1072
408 rAspIleAlaLysThrGlyGlyPhe.....ValGluThrS 420
1073 GAACGTTCAATTG.....TTAATGTTCTTTATTCGAG 1107
420 ergIleHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 436
1108 ACAGCAAGAGAACCT.....GTTTATCATGCTGACGAGTGGCT 1145
437 TrpLeuLysProAspAspValSerIleGluThrLeuThrSerGlyAr 453
1146 CAACAGTTATCGACCCAGACTGAATATGACAAATAATTTCTTAT 1194
453 gAsn.....AsnHnGlyLysAsnGlnGlyTyrHnT 464
1195GACAAAGCAAAAGTGAATTGATCTACTTACCAGC 1227
464 hGlyAspGlyThrLysGluSerProLysAsnSerIleSerLysPro 480
1228 AACATC...AACCAAGCGCGGCGGTGTGTTTGAAGGTAATTTAC 1274
481 ThrLeuHnAsnSerThrLeuGluGlnIleLeuAlaGlyGlySerTyrVa 497
1275 GGTCTGCTTAAACACAGAAAGTGGCAAGCGCGGCGCTTATATCA 1324
497 IAsnIleThrAlaAsnAsnArgIleTyrValAsnSerIleAsnLeuS 514
1325 GTGATGCACT...ACCGTTACTTGAAGTAAGGCGGTG...GCAAC 1368
514 eAsnGlySerLeuThrLeuHnIleHnIleAspAlaGlyLysIleAsn 530
1369 GACCGCTGTCAAAATCGCAAGGACGCTGTGTTCAAGCCAAAG 1418
531 GlyAspIleThrSerAsnGluAsnGlyAsnLeuThrIleLysAlaGlySe 547
1419 G.....GAAACCAAGGCTCGGTGCGGTGGGACGAGTAAAGTCATCT 1462
547 rTrpValAspValHisLysAsnIleThrLeuGlyThrGly.....PheL 562
1463 TAGATGACGAGCGAGAT.....CAAGCAAAA 1494
562 eAsnIleValAlaGlyAspSerValAlaPheGluArgGluGlyAspLys 578
1495 CAAGCTTAGTGAATTCGCTTGGTCAAGCGGACGAGCGGTCAACT 1544
579 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleHnVa 595
1545 GAATGCCGATTAATCAAGTTCACCCGACAAACTATTTGCGCTTGGC 1594
595 IAsnLysAspAspLys.....GlnHnArgPheAsn 606
1595 GCGGACGTTTGGATTTGAACGGCATTTGCTTCCACCGCATTCAA 1644
606 snValSerLeuAsnGlyHnGlyLysGlyLeuLysPheIleAlaAsnGln 622
1645 AATACGATGAAGGCGGATGATTCGAACCAATCAAGCAAAAGATC 1694
623 AsnAsn.....PheHnHisLysPheAspGlyGlu... 632
1695 CACCGTTACATTACAGGCAATAAAGATATTACTACAACCGGCAATACA 1744
633IleAsnIleSerGlyIleValIleHnIleAsnGlnHn 644
1745 ACAACTTGGATGCAAAAGAAATTTGCTCAACAGCTTGTGGTGGCAG 1794
645ThrLysLysAspValLysTyr.....TrpAsnAlaSer 655
1795 AAGATGCAACCAAAAGCAAGCGGCGCTCATCTGAAATTCACAAACGGA 1844

656 LysAspSerTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnIly 672
1845 AGAAGCGGATGGCACTTACTGCTTCCGGCGGACAAATTTAAAGCGCA 1894
672 sPheThrPheIleuysPheValAspSerGly..... 682
1895 ATATCAGCAAAACAAAGCAACTGTTTTCACGGCGAGACCGACCGC 1944
683SerAsnGlyIlnAspLeuArGSerSerArg..... 692
1945 CAGCGCTACATCATTTAGAGAGCGGCTGGTCAAATAATGAGAGTATCC 1994
692 692
1995 ACAAGAGAAATCGTGTGGGACACAGATGTGATCGACCGACATTTAAAG 2044
693ArgSerPheAlaG 697
2045 CGGAATACTTCATATTCAGCGCGGACAGAGCGGTTCGCCGCAATGTT 2094
697 LysValHisPheAsnGlyIleGlyIlySerThr..... 707
2095 GCGCAAGTGGAGCGGATGTGCTTAAAGCATACGCCACAGCATTTT 2144
708AsnPheAsnIleGlyAlaAsnAlaIlyAlaLeuPh 719
2145 CGGTGTGCGACCGCAT..... 2160
719 eLysLeuysProAsnAlaAlaThrAspProLysLysGluLeuProIleT 736
2161CAAAGCCACACATCTGTACAGCTTCGAGCTGAGCGGTGTGACA 2205
736 hrPheAsnAlaAsnIleThrAlaThrGlyAsnSerAspSerValMet 752
2206 AGTGTACCGCAAAACCATTCAGCAGATTAAGTATGATGCTTATTGAG 2255
753 PheAspIleHisAlaAsnLeuThrSerArgAlaIleGlyIleAsnMetAs 769
2256 CAAGACGCACTCAGAGCAATGTGAGCTTCGCGATCAGCAT..... 2301
769 pSerIleAsnIleThrGlyIleuAspPheSerIleThrSerHisAsnA 786
2302TTAATCTCACA 2313
786 rGAsnSerAsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThr 802
2314 GGA..... 2316
803 GlySerAsnPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyr 819
2317CTTGCCACACTCAACGCA 2335
819 rSerLysHisAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyAla 836
2336 ATCTTAGTGCAGCGGAGAGACAGCATATACGTTACGCGCAACGCCAC 2385
836 snValThrLeuGlyGlyGlu.....AsnSerSerSerSer 847
2386 CAAAACGCGCAACCTCAGCTCGTGGGCAATGCCAAGACATTTAATCA 2435
848 IleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThrLeuGluAl 864
2436 AGCCACATTAAACGGCAACATCGGCTTCGACATGCTTCAATTAATC 2485
864 aAspThrSerAsnSerAsnThrGlyLeuLys...LysArgThrLeuThrI 880
2486 TAAGCAACACAGCGGTACAAAAGCGCATGTGACGCTTCGCAACAGCT 2535
880 eugLysAsnIleSerValGlu...GlyAsnLeuSerLeuThr...GlyAla 894
2536 AAGGCAACGTAAAGCATTCGCACTCAACATGCTCCCTAGCGCA 2585
895 AsnAlaAsnIleVal.....GlyAsnLeuSerIleAlaGlu 906
2586 TAAGCAGTATTCATTGTA.....ACAGCCGCTTACCGGAA 2626
906 uAspSerThrPheLysGlyGluAlaSerAspAsnLeuAsnIleThrGlyT 923
2627 AATACAGCGCGCAAGATACGCGATTACCTTAAAGACAGCGCAATGG 2676
923 hrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLys..... 935
2677 ACGCTCCGTCGCGACGAAATTAGCAATTTAAACCTTGAACAGCGCAC 2726
936GlyValValLysLeuGlyAspIleAsn..... 944
2727 CATTAACATCAATTCGCTATCGACAGATGCGGACGCGCAACCG 2776
944 944
2777 GCAGTGGCAGATGCGCGCGCGCGCTTCGCGGCTTCCTATTATCC 2826
945AsnLysGlyIleuAsn 950
2827 GTTACGCGCGCAACTTCGCGCAGATCCGTTTCAACAGCGTACGCTAA 2876
951 IleThr.....ThrAsnAlaSerGlyThrGlnLysThrIle...IleAs 964
2877 CGGCAAAATG...AACGTCAGAGAACATTCGCTTATGTG..... 2916
964 ngLysAsnIleThrAsnGlyLysGlyAspLeuAsnIleLysAsnIleLysA 981
2917GACTCTTCGCTACCGCAGCGGCAATTTGAAGCTG 2952
981 laSpAlaGluIleGlnIleGlyIlyAsnIleSerGluLys..... 994
2953 GCGGAAGTTCCGAAGCACTTACACTTGCT.....GTCAACA 2993
995GluLysAsnLeuThrIleSerSerAspLysValAsnI 1007
2994 TACCGGCAACGAACCCGTAGTCTCGAGCAATTCAGCTA.....G 3034
1007 eThr.....AsnGlnIleThrIleLysAlaGly 1017
3035 TGGA...GGAAGACACACACCGCTTCGCAAAATTTAATTCACC 3081
1017 alGluIlyLysArgSerAspSerSerGluAlaLysAlaAsnLeuThr 1033
3082 CTCGAAACGAAACACGTCGATGCGCGCATGCGCTTATCAGCTTATCCG 3131
1034 IleGln..... 1035
3132 CAAAGACGCGAGTTCGCTGCATATATCCGTCAAGACAGCTTT 3181
1036ThrLysGluLeuLysLeuA 1042
3182 CCGCAAACTCGGCAAGCGGCGGAAACAGAGGCGCTTGACGCGAANA 3231
1042 laGlyAspLeuAsnIleSerGlyPheAsnLysAlaIleThrAlaLys 1058
3232 CAGGCACACTTGCCTCCAAACACAGCGGGAAGAACAGACCGCAAG 3281
1059 AsnGlySerAspLeuThrIleGlyAsnAlaSerGlyAlaAsnAlaSpAl 1075
3282 C.....CTTGACGCGCTGATTCGGGCC...G 3304
1075 alLysLysValThrPheAspLysValLysAspSerLysIleSerThrAspG 1092
3305 GCGCAATGCGACCGAAAGCAGAA...AGTGTTCGCAACGCGCGCG 3351
1092 LysHisAsnValThrLeuAsnSerGluValLysThrSerAsnGlySerSer 1108
3352 CAGCAGCGCGGGAATATGCG...GGCATTTACAGCGCGAGAGAACG... 3396
1109 AsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAlaLysAspVa 1125

```

3397 .....AAAAAGGGTGACGGCGATTAAG 3421
1125 LThrValAsnAsnValThrSerHisLysThrIleAsnIleSerAla 1142
3422 ACACCGCTTGGCGAAGACGGCGAAGCGGACCGCGCTACAC 3471
1142 LaAlaGlyAsnValThrThrLysGluGlyThrIleAsnAlaThrThr 1158
3472 GCCTTCCCGCCGCGCGCGCGGATTTGGCGCAACCGCACCC 3521
1159 Gly.SerValGluValThrAlaGlnAsnGlyThrIleLysGlyAsnIle 1175
3522 CCAACCGCAACCCCAACCGCAGCGCGATCATCAGCGCTATGCCATA 3571
1175 hrSerGlnAsnValThrValThrAlaThrGlnAsnLeuValThrThrGlu 1191
3572 GCGGTTGAGTATTTCCGCCCAACCGCTTTCGCGCTTCCGCTACAG 3621
1192 AsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer..... 1205
3622 GACGAATTTGACGCGCTGTTTCCGCAACCGCGCACCGCTTGA. 3670
1206 .....ThrLysThrGlyAspIleLysGlyLysI 1215
3671 ..CAAGCGCATTCGGGACACCAACACTACCGTTCG..... 3706
1215 LegLysThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThrLeu 1231
3707 AGATTTCCGCGCTACCGCAACCAACCGCGCAATTCGCGATG 3756
1232 LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAspAlaG 1248
3757 CAGAAAACCTCGGACGCGCGCGCTGCGATCTGTTTTCGCAACCG 3806
1248 ValAlaLeuThrThrAlaGlySerThrIleSerAla.....ThrThrG 1263
3807 GA..... 3808
1263 LysAlaAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1279
3809 CCGGAACACCTTCGACGACGCGATCGCAACTCGGACGCGTTGGCCAC 3858
1280 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAla 1296
3859 GGTGCGCTTTTGGGCAAT.....ACGCGATCGGACGTTTGA 3896
1296 IGIAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLys 1313
3897 CATCGGATCAGCGCGCGCGGTTTACTAGCGGACGCTTTCAGACG 3946
1313 eutThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1329
3947 GCATCAGAGCAAAA.....TCCGCGCGCGCGTGTG 3978
1330 SerSerSerGlySerValThrLeuGlyThrIleSerGlyAsnThrValAs 1346
3979 CATTCAGCGCATTAAGCGAAGAT.....ACGCGCGACGTTT 4013
1346 nValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysValG 1363
4014 CCGGCGATTGCGCATCGAACCACATCGCGCGCAACGCGCTATTTGCTC 4063
1363 LuAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu.....Ser 1375
4064 AAAAAGCGGATTCAGTACGAAGAAGTCATATCGCACCGCGCGCTT 4113
1376 GlyLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGly.. 1391
4114 GCATTCACCGCTACCGCGCG 4135
1392 .GlnThrThrLeuThrAlaLys 1398
seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-728-470-10

```

```

seq_documentation_block:
/ Sequence 10, Application US/08728470
/ Patent No. 5928651
/ GENERAL INFORMATION:
/ APPLICANT: Barenkamp, Stephen J
/ TITLE OF INVENTION: High Molecular Weight Surface Proteins
/ TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Shoemaker and Matlare, Ltd.
/ STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
/ CITY: Arlington
/ STATE: Virginia
/ COUNTRY: U.S.A.
/ ZIP: 22202-0286
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/728,470
/ FILING DATE:
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/302,832
/ FILING DATE: 16-MAR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US PCT/US93/02166
/ FILING DATE: 16-MAR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 9205704.1
/ FILING DATE: 16-MAR-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Berkstresser, Jerry W
/ REGISTRATION NUMBER: 22,651
/ REFERENCE/DOCKET NUMBER: 1038-633
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 415-0810
/ TELEFAX: (703) 415-0813
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1529 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-728-470-10

alignment_scores:
Quality: 292.50 Length: 1494
Ratio: 0.415 Gaps: 77
Percent Similarity: 47.189 Percent Identity: 20.415

alignment_block:
US-09-303-518D-653 x US-08-728-470-10 ..

Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529

178 AATAAGCAAGTTGTCAGTCGGGCGCAAGATATTTAGTTTACMACAA 227
111 111 .....111111 111111 .....111
127 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 139
228 AAAAGGAGTTGTGCGCAATCGATGACGAAGAAGCCCGCATGATTGATT 277
111 111 .....111 .....111
140 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 149
278 TTTCTGTGTATCGCGTAAGGC.....GTGCGGCGATTGCGGCGGAT 321
111 111 .....111 .....111
150 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 165

```

```

322 CAA.....TATATGTGAG 335
166 LysIleThrIleSerAspIleLeuAsnProThrIleThrTyrSerIleLeu 182
336 CGTGGCAGCATTAACGGCGCTATTAACAATGTTGATTTGGTCGAGAGGAA 385
182 aaIaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaIleGlyG 199
386 GCAATCCCGATCAGCACCGCTTTCTTACCAAAATGTGAAGAATAAT 435
199 LysAsn..... 200
436 TATAAACGAGGAGTAACGGCCATCCTTATGGGGGATTAATATATGCC 485
200 ..... 200
486 GCGTTTCACAAATTTGTACAGATGCAGAACTGTGAGATGACAGATT 535
200 ..... 200
536 ATATGATGGGTGAATACCGTGATTTAAATAATACCTGATCGTGT 585
201 .....IleAsn.....Val 203
586 CGAATCGAGACGACAGACAATATTGGCGGTGATGAGACGACGACCA 635
204 ArgAlaIaIaThrIleArgAsnLysGlyLysLeuSerAlaAspSerValSe 220
636 TAACCGGGAAGTTCAATATCATATTGCGAACGCAATATCTTGG.....C 679
220 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGlyLysIleVala 237
680 TCGTGGGTGCAATACCTTTCACAAATATGATCAGCT...GGTGGACACA 726
237 IuIleGlyGlyValIleSerAlaIleAsnGlnAlaLysGlyGlyLys 253
727 GTCAACTAGTAGGCAAAATAATTAACATAGCCCATATGTTTATACC 776
254 LeuMetIleThrGlyAspLysValThr.....LeuLys 264
777 AACGAGAGGCTCATTTGGCGACAGCTGCTCCACATGTTATATGATG 826
264 sThrGlyValValIleAspLeuSerGlyLys..... 274
827 CCCAAAGCAAAAGTGTATTAATGAGGATTAACAGCAAGCAACCC 876
275 .....GluGlyGlyGluThr 279
877 TATATAGGA.....AAAGCAATGGCTTCACAGTAGT 908
280 TyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 296
909 TCGTAAGATGTGTTCTATGATGAATCTTTCGAGATACCATTCAG 958
296 aLysLys.....ThrThrLeuGluLysGlySerThrIleAsnV 309
959 TATCTACGACACACATCAATAATGGAAA..... 987
309 al.....SerGlyLysGlyLysGlyArgAlaIle 319
988 .....TACTTTTAAACGACATATAATATGGCGACGAGAA 1022
320 ValTrpGlyAspIleAlaLeuIleAspLysAsnIleAsnAlaGlnGlySe 336
1023 AATGATGCCAAACATAACACTATTCCTCTACCTTATAGATTAAACAC 1072
336 rAspIleAlaLysThrGlyGlyPhe.....ValGluThrS 348
1073 GAACCGTTCAATG.....TTTAATGTTTCTTTATCCGAG 1107
348 ergLysHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 364
1108 ACAGCAAGAGAACT.....GTTTATATGCTGACAGGTGGGGT 1145

```

```

365 TrpLeuLeuAspProAspAspValSerIleGluThrThrLeuThrSerGlyAr 381
1146 CAACAGTTATGACCCACAGCTGAATATGAGAAATAATTTCTTTAT. 1194
381 gAsn.....AsnThrGlyGluAsnGlnIleYrThrT 392
1195 .....GACAAAGGAAAGGTGAATGATCTACTTAACAGC 1227
392 hGlyAspGlyThrLysGluSerProLysGlyAsnSerIleSerLysPro 408
1228 AACATC...AACCAAGCGCGCGGCTTTGTATTTTGAAGGTAATTTTAC 1274
409 ThrLeuThrAsnSerThrLeuGluGlnIleLeuAlaArgArgLysSerYVa 425
1275 GGTCTCGCTTAAACAAACAGCAACGTCGGCAAGCGCGGCGCTCATATCA 1324
425 lAsnIleThrAlaAsnAsnArgIleTyrValAsnSerIleAsnLeuS 442
1325 GGTATGGCAGT...ACGTTACTTGAAGTAACGCGGTG...GCAATC 1368
442 eAsnGlySerLeuThrLeuHisThrLysArgAspGlyValLysIleAsn 458
1369 GACCGCTGTCCAAATCGGCAAGGCAACGCTGCTGTTCAAGCCAAAG 1418
459 GlyAspIleThrSerAsnGlnLysAsnIleThrLeuLysAlaGlySe 475
1419 G.....GAAACCAAGCGCTCGTCAGCGTGGCGAGGTAAGCAACT 1462
475 rTrpValAspValHisLysAsnIleThrLeuGlyThrLys.....PheL 490
1463 TAGATCAGCAGCGGAGCAGT.....CAAGGCAAAAA 1494
490 euAsnIleValAlaGlyAspSerValAlaPheGluArgGluLysLys 506
1495 CAAGCCTTAGTGAATCGCTTGTGTCAGCGGCGAGGAGCGTGCAACT 1544
507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThVa 523
1545 GAATGCCGATTAATCAAGTCAACCCGACAAACTATTTGGCTTTCGCG 1594
523 lAsnLysAspAspLys.....GlnPheArgPheAsnA 534
1595 GCGGAGCTTGGATTGAACGGGCAATGCTTTCCTCCACCGCACTTAA 1644
534 snValSerIleAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 550
1645 AATACCGATGAAGGGCGATGATGTCAACCAACATCAAGACAAAGATC 1694
551 AsnAsn.....PheThrHisLysPheAspGlyLys.. 560
1695 CACCGTTACATTAACAGCAATTAAGATTAATTAACACCGGCAATTAACA 1744
561 ...LeuAsnIleSerGlyIleValThrIleAsnGlnThr..... 572
1745 ACAACTGGATAGCAAAAAAGAAATTTGCTACACAGGTTGTTGGCAG 1794
573 .....ThrLysLysAspValLysTyr.....TrpAsnAlaSer 583
1795 AAGATGCAACCAAAAGCAAGGCGGCTCAATCTG..... 1830
584 LysAspSerTyrTrpAsnValSerSerLeuThrLeuAlaSerThrValGlnLys 600
1831 .....AATTACCAACCGGAAG 1846
600 sPheThrPheIleLysPheValAspSerGlySerAsnSerGlnAspLeuA 617
1847 AAGCGATGCGACTTACTG.....CTTCCGCC 1875
617 rSerSerArgSerPheAlaGlyValHisPheAsnGlyIleGlyGly 633
1876 GGAACAAATTTAAACGCAATATACGCAACAAACGCAAACTGTTT 1925

```

634 LysThrAsnPhasn.....IleGlyAlaAsnAlaLysAlaLeuph 647
 1926 CAGCGGACGACGACGA.....CCGACG 1948
 647 elysleuLysProksnAlaIatThrAspProLysLysGluLeuProLiet 664
 1949 CCTACCAATTCATTTAGGAAGCGGGTGTCAAAATATGAAAGGTATCCACAA 1998
 664 hrPhasn.....AlasnIleThrAlaThrGlyAsnSerAsp 676
 1999 GGAAATTCGTGTGGACACAGATTCATGACCGCATTTAAAGC.. 2046
 677 SerSerValMetPhasnIleHis..AlasnIleThrSerAlaAla 692
 2047GAAACTTCATTTCAAGCGGACAAAGCGGTGTTT 2083
 692 agLYIleAsnMetAspSerIleAsnIleThrGlyLeu..... 705
 2084 CCGCCAAATGTTGCCAAAGTGGAAAGCGATTTAGCAATTCAGGCC 2133
 706AspPheSerIleThrSerHisAsn 713
 2134 CAACGAGTTTCGGTGTCCGACCCATCAAAAGCCACACATTCGTACAGC 2183
 714 Arg.....AsnSerAsnAlaPheGluIle 722
 2184 TTGGACTGGACGGGTCTGACCAAGTTGTACCGAAATTCATTTA 2227
 722 slyAspLeuThrIleAsnAlaThrGlySerAsnPheserLeuLysGlnT 739
 2228 CCGACGATAAGTATGCTTCATTTAGCAAGACGACATCAGAGCAAT 2277
 739 hrLysAspSerPheTyraSnGlnLysSerLys..... 749
 2278 GTACGCTTCCGATCAGCTCATTTAAATCTACAGAGCTTGCACACT 2327
 750HisAlaIleAsnSerSerHisAsnLeuThrIle 761
 2328 CAACGCAATCTTAGTCAGCGGAGACAGCACTATACGGTTACGGCA 2377
 761 uGlyGlyAsnValThrLeuGlyGlyGlu.....AsnSer 773
 2378 ACGCCACCAAAACGCAACCTCAGCTCGTGGCAATGCCCAAGCAACA 2427
 773 erSerSerIleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThr 789
 2428 TTATATCAAGCCATTTAAACGGACACATCGGCTTGGACATGCTTC 2477
 790 LeuGlnAlaAspThrSerAsnSerAsnThrGlyLeuLys..LysArgTh 805
 2478 ATTAAATCTAGCAACAACGCCGTACAAACGGAGTGTGACGCTTCCG 2527
 805 rLeuThrLeuGlyAsnIleSerValGlu..GlyAsnLeuSerLeuThr 820
 2528 ACAACGCTAAGCAACGATTAAGCATTCGACACTCAACGCAATGCTTC 2577
 821 ..GlyAlaAsnAlaAsnIleVal.....GlyAsnLeuSer 831
 2578 CTAGCCGATTAAGCAGATTTTCATTTGAA.....AACACGCCGCTT 2618
 832 IleAlaGluAspSerThrPheLysGlyAlaIleSerAspAsnLeuAsn 848
 2619 TACCGGAAATATCAGCGGCGGCAAGATACGCACTTAAAGACA 2668
 848 ethrGlyThrPheThrAsn..... 854
 2669 GCGAATGAGACTCCGTCGCGACGCAATTAAGCAATTTAACTGTAC 2718
 855AsnGlyThr.....AlaAsnIleAsnIleLys 863
 2719 AAGCCCACTTACACTTCCGCTATGACAGACATGCGGACGGCGC 2768
 864 GlnGlyValValLysLeuGlnGlyAspIleAsnAsnLysGlyGly..... 878

2769 GCAAAACGGGACATGCGCGAGATCCGGCGCGCCCTTGGCCCGCTTCC 2818
 878 878
 2819 TATTATCCGTTACGCGCAACTTGCAGAAATCCGTTTCAACAGCGTG 2868
 879 ..LeuAsnIleThr.....ThrAsnAlaSerGlyThrGlnLysThrIle 892
 2869 ACCGTTAACGGCAATTTG...AACGTCAGGAAACATTCGCTTATGTC 2915
 893 ..IleAsnGlyAsnIleThrAsnGlnLysGlyAspLeuAsnIleLysAs 908
 2916 G.....GAACCTTCGCTTACCGCAGCGCAAT 2944
 908 nIleLysAlaAspAlaGluIleGlnIleGlyAlaAsnIleSerGlnLys 924
 2945 TGAAGCTGGCGGAAAGTTCCGAAGCACTTACACTTGGCT..... 2985
 925GluGlyAsnLeuThrIleSerSerAspLys 934
 2986 GTACCAATTAACCGGACAGAACCCGTAGTCTGACGCAATTAAGCGTA.. 3033
 935 ValAsnIleThr.....AsnGlnIleThrIleLys 944
 3034GTGGAA..GGAAGACACACACCGCTGTCCGAAATCTTA 3073
 944 salagIyValGluGlyGlyArgSerAspSerSerGlnAlaGluAsnAla 961
 3074 ATTTCACCTTGCACAAAGCAACGTCGATCCGCGCATGGCTTATCAG 3123
 961 snLeuThrIleGln..... 965
 3124 CTATTCGCAAAAGACGGCGAGTTCGCGCTGATTAATCCGTTCAAGACA 3173
 966ThrLysGluLe 969
 3174 AGAGCTTTCGCAACAACTCGGCAAGCGGAGAAACAGAGCGGCTTGA 3223
 969 uLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleT 986
 3224 CCGCAAAACAGGACACACTTGCCTCAACACAGCGGCAAAAGACAAC 3273
 986 hrAlaLysAsnGlnLysAspLeuThrIleGlyAsnAlaSerGlyLysAsn 1002
 3274 GCGCAAGC.....CTTGACGCGCTGATTCG 3299
 1003 AlaAspAlaLysLysValThrPheAspLysValLysAspSerLysIleSe 1019
 3300 GGCC...GGCGCAATGCCACGGAAGGACAGA...AGTGTGCCGAC 3343
 1019 rThrAspGlyHisAsnValThrLeuAsnSerGlnValLysThrSerAsn 1036
 3344 CCGCCGCGGACGAGCGGCGGAAATGCG...GGCATTAAGCAGCGGAG 3390
 1036 LysSerSerAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAla 1052
 3391 GAAGAG.....AAAAACGGGTGACGC 3413
 1053 LysAspValThrValAsnAsnAsnValThrSerHisLysThrIleAsnI 1069
 3414 GGATTAAGACACCGCTTGGCAAAACAGCGGCAACCGGCGCGG 3463
 1069 eSerAlaAlaAlaGlyAsnValThrThrLysGlnGlyThrThrIleAsn 1086
 3464 CTACACGCGCTTCCCGCGCGCGCGCGCGGATTTGCGGAA 3513
 1086 IeThrThrGly..SerValGluValThrAlaGlnAsnGlyThrIleLysG 1102
 3514 CCGCAACCCCAACCGCAACCCGACGCGACGATGATCAGCGGTTA 3563
 1102 ysnIleThrSerGlnAsnValThrValThrAlaThrGlnLysLeuValT 1119

```

3564 TGCCAAATGACGGTTGATGATTTTCGCGCAGCTCAACAGCGTTTCG 3613
      :::::|||||
1119 hrThngluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135
      :::::|||||
3614 CCGTACAGACGAAATGGACCGCGTGTTCGCGAAGCCGCCGCAACGCC 3663
      :::::|||||
1136 .....ThrystrhglyAspilely 1142
3664 GTTTGA...CAAGGGCATCGGGGAGACCAACACTACCGTTCCG... 3706
      ||| :::::|||||
1142 sglYelylegIuserThrSerGlyAsnValAsnIleThrAlaSerGlyA 1159
      ||| :::::|||||
3707 .....AGATTTCGCGCGCTACCGCCCAACAAACGCGCGCAAA 3748
      ||| :::::|||||
1159 snThrleuLysValSerAsnIleThrGlyGlnAspValThrValThrAla 1175
      ||| :::::|||||
3749 TCGGTATGCAAAAACCTCGGACGCGCGCGCTCGCATCTCTTTTCG 3798
      :::::|||||
1176 AspAlaGlyAlaLeuThrThrAlaGlySerThrIleSerAla.... 1190
3799 CACAAACCGGA..... 3808
      ||| :::::|||||
1191 .ThrThrGlyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyL 1207
3809 .....CCGAAACACCTTCGACGACGCGCATCGCACTGCGCACGCC 3850
      :::::|||||
1207 yValIgluserSerSerGlySerValThrleuValAlaThrGlyAlaThr 1223
3851 TTGCCACAGGTGCGCTTTTCGGCAAT.....ACGCATCGGC 3888
      ||| :::::|||||
1224 leuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSe 1240
3889 AGGTTGCATCGCATCGCATCGCGCGCGGCTTTAGTAGCGCGACGCT 3938
      ||| :::::|||||
1240 rGlyLysleuThrSerThrValGlySerThrIleAsnGlyThrAsnSerV 1257
3939 TTCAGACGCGCATCGACGCCAAA.....TCCGCCGCC 3970
      ||| :::::|||||
1257 alThrThrSerSerGlnSerGlyAspIleGlyThrIleSerGlyAsn 1273
3971 GCGTGTGCTGATTCAGCGCATTCAGCGCAAGT.....ACGCG 4005
      ||| :::::|||||
1274 ThrValAsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290
4006 GCAGGTTTCGCGCGATTCGCGCATCGCAACCGCATCGCGCGCAACGCGTA 4055
      ||| :::::|||||
1290 alYValGluAlaLysAsnGlyAlaAlaThrleuThrAlaGlu..... 1304
4056 TTTCTCCAAAAAGCGGATACCGATACGAAAGCTCAATATCGCCACCC 4105
      ||| ||| :::::|||||
1305 .....SerGlyLysleuThrThrGlnThrGlySerSerIleThrSer 1319
4106 CGGGCCTTGATTCACCGCTACCGCGCGG 4135
      :::::|||||
1320 AsnGly...GlnThrThrleuThrAlaLys 1328
seq_name: /cgn2.6/ploddata/1/iaa/bb_COMB.pep:us-08-719-641-10
seq_documentation_block:
: Sequence 10. Application US/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.

```

```

: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1529 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-719-641-10

alignment_scores:
  Quality: 292.50      Length: 1494
  Ratio: 0.415         Gaps: 77
  Percent Similarity: 47.189   Percent Identity: 20.415

alignment_block:
US-09-303-518D-653 x US-08-719-641-10 ..
Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529

178 AATAAGGCAAGTTTCAGTCGCGGCGCAAGATATTGAGTTTACAACA 227
    ||| ||| :::::||||| ||||| :::::|||||
127 AsnHleGlyleuThrValGly...LysAspGlySerVal..... 139
228 AAAAGGGAGTTGTGCGCAATCGATGACGAAAGCCCGCATGATGAT 277
    :::::||||| :::::|||||
140 .....AsnleuIleGlyGlyValLysAsnGlu..... 149
278 TTTCTGTGATTCGCGTAAACGCC.....GTGCGCATTTGGCGGCGAT 321
    :::::||||| ||||| :::::|||||
150 .GlyValIleSerValAsnGlyGlySerIleSerleuLeuAlaGlyGln 165
322 CAA.....TATATTGTGAG 335
    :::::|||||
166 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 182
336 CGTGCACATACGCGCGCTATACAAATGTTGTTGGTGGGAGGAA 385
    :::::||||| :::::|||||
182 aaIaProGluAsnGluAlaIleAsnleuGlyAspIlePheAlaLysGly 199
386 GCAATCCCGATGACACCGCTTTCTTACCAATGTGAAAGAATATAT 435
    :::::|||||
199 LysAsn..... 200
436 TATTAAGCAGGACTAAGCGCATCTTATGCGCGGATTAATCATATGCC 485
200 ..... 200

```



```

692 aglyleasmetaspserileasnilethrglyleu..... 705
2084 CCGCAATGTTCAGCAAGTGGAGGCGCATTTAGCAATCAGGCC 2133
706 .....AspHisSerIleThrSerHisAsn 713
2134 CAGCAGTTTTCGGTGTCCGACCCCAAGCAACAAATCTGACAC 2183
714 Arg.....AsnSerAsnAlaPheGluIleLeu 722
2184 TTCGGACGACGGGTCTGACAGTTGACGAAAAACATTT.....A 2227
722 slyspserleuthrileasnialthrGlySerAsnPheserleuysGlnT 739
2228 CCGAGATAAAGTATGCTTCATTGACGACGACGACATCAGAGCAAT 2277
739 hrlYsAspSerPheTyrAsnGlnIuTyrSerIys..... 749
2278 GTCAGCCTTGGCCGATCAGCTCATTTAAATCTACAGGACTTGGCACA 2327
750 .....HisAlaIleAsnSerSerHisAsnLeuThrIleLe 761
2328 CAAGCGCATCTTAGTCAGGCGGAGACAGCCATATACGTTACGGCA 2377
761 uGlyGlyAsnValThrLeuGlyGlyGlu.....AsnSer 773
2378 ACAGCAGCAAAACGCAACCTCAGCTGTGGCAATGCCAAGCAACA 2427
773 erSerSerIleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThr 789
2428 TTTATATCAGGCACATTAAAGCGACACATCGCTTCGACAAATGCTTC 2477
790 LeuGlnAlaAspThrSerAsnSerAsnThrGlyLeuLys...LysArgTh 805
2478 ATTTATCTATGACAAACAGCGCTGACAAAGCGAGCTGACGCTTCG 2527
805 rleuthrleuGlyAsnIleSerValGlu...GlyAsnLeuSerleuthr. 820
2528 ACAACGCTAAGGCAACGATAGCCATTCGCACTCAACGCAATGTCTCC 2577
821 ..GlyAlaAsnAlaAsnIleVal.....GlyAsnLeuSer 831
2578 CTAGCCGATAGGCGATATTCATTTGA.....AACAGCGCTT 2618
832 IleAlaGlnAspSerThrPheLysGlyAlaLaserAspAsnLeuAsnI 848
2619 TACCGAAAAATATCAGCGCGCAAGATACGCTTAAAGACA 2668
848 eThrGlyThrPheThrAsn..... 854
2669 GCGAATGAGCGCTGCCGTGGGCAAGCAATTAAGCAATTAACCTTGAC 2718
855 .....AsnGlyThr.....AlaAsnIleAsnIleLys 863
2719 AACGCCACCATTTACACTTCGCTATCGACAGCATGCGGAGGCGC 2768
864 GlnIleValValLysLeuGlnGlyAspIleAsnAsnLysGlyGly.... 878
2769 GCAAAACGGGAGTGGCGAGATGGCGCGCCGCCGCTTCC 2818
878 ..... 878
2819 TATTATCGTTAGCCCGCAACTCGGCAATCCGTTTCAACAGCGTG 2868
879 ..LeuAsnIleThr.....ThrAsnAlaSerGlyThrGlnLysThrIle 892
2869 ACGGTAAACGGCAATTTG...AACGGTCAGGAGACATTCGCTTATGTC 2915
893 .....IleAsnGlyAsnIleThrAsnGlnLysGlyAspLeuAsnIleLysAs 908
2916 G.....GACTCTTCGGCTACCGGAGGCGCAAT 2944
908 nIleLysAlaAspAlaGlnIleGlnIleGlyGlyAsnIleSerGlnLys. 924

2945 TGAAGCTGGGGAAAGTTCCGAAGGCACTTACACCTTGCT..... 2985
925 .....GlnGlyAsnLeuThrIleSerSerAspLys 934
2986 GTCACAAATACCGGCAACGACACCCGTAAGTCTGACGCAATTCAGGTA.. 3033
935 ValAsnIleThr.....AsnGlnIleThrIleLeu 944
3034 .....GTGAA..GGAAGACACACACCGCTGTCCGAAATCTTA 3073
944 salaglyValGlnGlyGlyArgSerAspSerSerGlnAlaGlnAlaAla 961
3074 ATTACACCTGCAAAACGAAACAGTCGATGCGGCGCATGCGTTATCAG 3123
961 snleuthrIleGln..... 965
3124 CTATCCGCAAAAGACGCGAGTTCGCTGCATTAATCCGTCAGAAACAA 3173
966 .....ThrLysGluLeu 969
3174 AGAGCTTTCGCAAACTCGGCAAGGCGGAGAAACAGAGCCGCTTGA 3223
969 uLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGlnIleT 986
3224 CCGCAAAACAGGACACAACTTCCGCCAAACAGAGCGGAAAAAGACAC 3273
986 hValAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyLysn 1002
3274 GCGCAAGC.....CTTGACGCGCTGATTCG 3299
1003 AlaAspAlaLysLysValThrPheAspLysValLysAspSerLysIle 1019
3300 GGCG...GGCGCAATGCCACCGCAAAAGGCAAGA...AGTGTGCGGAC 3343
1019 rThrAspGlyHisAsnValThrLeuAsnSerGlnValLysThrSerAsnG 1036
3344 CCGCCGCGGAGGAGCGGGGAAAAATGCG...GGCATTATGACGAGCGAG 3390
1036 LysSerSerAsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAla 1052
3391 GAAGAG.....AAAAACGGGTCAGCGC 3413
1053 LysAspValThrValAsnAsnAsnValThrSerHisLysThrIleAsnI 1069
3414 GGATTAAGACACCGCTTGGCAAAACAGCGGCAACCGGACCGCGCG 3463
1069 eserAlaAlaAlaGlyAsnValThrThrLysGlnGlyThrThrIleAsnA 1086
3464 CTACACCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCGCGCA 3513
1086 lAthrThrGly..SerValGlnValThrAlaGlnAsnGlyThrIleLysG 1102
3514 CCGCAGCCCAACCGCAACCCCAACCGGAGCGGCACTGATCAGCGTTA 3563
1102 yAsnIleThrSerGlnAsnValThrValThrAlaThrLysLeuValT 1119
3564 TCCCAATAGCGGTTGAGTGAATTTTCGCGCACGCTCAACAGCGTTTCG 3613
1119 hThrGlnAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135
3614 CCGTACAGGAGCAATGAGACCGCGTGTTCGCAAGACCGCGCAACGCC 3663
1136 .....ThrLysThrGlyAspIleLeu 1142
3664 GTTTGA...CAAGCGCATCCGAGGACCAACAACTACGCTTGC... 3706
1142 sGlyGlyIleGlnSerThrSerGlyAsnValAsnIleThrAlaSerGlyA 1159
3707 .....AGATTTCGGCGGCTACCGGCAACAAACGACCTGGGCAAA 3748
1159 snThrLeuLysValSerAsnIleThrGlyGlnAspValThrValThrAla 1175

```



```

3749 TCGGATGACAGAAAAACCTCGGACGGCGCGTCGCGCATCCTGTTTCG 3798
      : : : : : | : : : : : | : : : : : | : : : : :
1176 AspalagialaleuthrThrThralaGlySerThrIleSera1a..... 1190
      : : : : : | : : : : : | : : : : : | : : : : :
3799 CACAACCGGA..... 3808
      : : : : : | : : : : : | : : : : : | : : : : :
1191 .ThrThrGlySna1aAen1leThrThrLyThrGlyAsp1leasnGlyL 1207
      : : : : : | : : : : : | : : : : : | : : : : :
3809 ..... CCGGAACACCTTCGACGACGATCGGACACCTCGGACGGC 3850
      : : : : : | : : : : : | : : : : : | : : : : :
1207 ysva1G1userSerSerGlySerValThrLeuValAlaThrGlyAlaThr 1223
      : : : : : | : : : : : | : : : : : | : : : : :
3851 TTGCCACAGGTGCGCTTTTCGGCAAT.....ACGGATCGGC 3888
      : : : : : | : : : : : | : : : : : | : : : : :
1224 LeuAlaValGlySna1leSerGlySnaThrValThrIleThrAlaAspSe 1240
      : : : : : | : : : : : | : : : : : | : : : : :
3889 AGGTTGCATTCGCGCATCGGCGCGCGGCTTTAGTAGCGGACGCT 3938
      : : : : : | : : : : : | : : : : : | : : : : :
1240 rGlyLySerThrSerThrValGlySerThrIleasnGlyThrsnSerV 1257
      : : : : : | : : : : : | : : : : : | : : : : :
3939 TTCAGCGGCATTCAGAGGCAAA.....TCGGCGGC 3970
      : : : : : | : : : : : | : : : : : | : : : : :
1237 alThrThrSerSerGlnSerGlyAsp1leGlyThrIleSerGlyAsn 1273
      : : : : : | : : : : : | : : : : : | : : : : :
3971 GCGTCTGCTATTCAGCATTCAGGCAAGAT.....ACGCG 4005
      : : : : : | : : : : : | : : : : : | : : : : :
1274 ThrValaSnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290
      : : : : : | : : : : : | : : : : : | : : : : :
4006 GCAGGTTTCGCGCATTCGCGCATCGAACCCGACATCGGCGGACGCGCTA 4055
      : : : : : | : : : : : | : : : : : | : : : : :
1290 aLyValG1uAlaLySnaGlyAlaAlaThrLeuThrAlaGlu..... 1304
      : : : : : | : : : : : | : : : : : | : : : : :
4056 TTTCGTCGCAAAAGCGGATTCAGGATACGAAAGTCATATTCGCCACCC 4105
      : : : : : | : : : : : | : : : : : | : : : : :
1305 ....SerGlyLySerLeuThrThrGlnThrGlySerSerIleThrSer 1319
      : : : : : | : : : : : | : : : : : | : : : : :
1320 AsnGly...GlnThrThrLeuThrAlaLyS 1328
      : : : : : | : : : : : | : : : : : | : : : : :
seq_name: /cgn2_6/ptodata1/1aa/bb_COMB.pep:US-08-460-269C-6

seq_documentation_block:
: Sequence 6, Application US/08460269C
: Patent No. 6197548
: GENERAL INFORMATION:
: APPLICANT: CLARE, JEFFREY J.
: ROMANOS, MICHAEL A.
: TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
: YEAST
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
: STREET: 2200 Clarendon Blvd., Suite 1400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,269C
: FILING DATE: 02-Jun-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Lebovitz, Richard M.
: REGISTRATION NUMBER: 37,067
: REFERENCE/DOCKET NUMBER: Popov-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 243-6333
: TELEFAX: (703) 243-6410

```

```

: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 922 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6

alignment_scores:
      Quality: 275.50      Length: 1162
      Ratio: 0.566      Gaps: 54
Percent Similarity: 41.910      Percent Identity: 21.687

alignment_block:
US-09-303-518d-653 x US-08-460-269C-6 ..

Align seq 1/1 to: US-08-460-269C-6 from: 1 to: 922

1366 AGACGCGCCTGTCGCAAAATCGGCAAGG.....ACGCT 1400
      : : : : : | : : : : : | : : : : : | : : : : :
2 AsnMetSerLeuSerArgIleValLysAlaAlaProLeuArgArgThr 18
1401 GCGTGTCAAGCCAAAGG..... 1419
      : : : : : | : : : : : | : : : : : | : : : : :
18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaAlaProAlaIleTyAlaA 35
1420 ....GAAACCAAGCGTCGCGTCAGCGTGCGGACGCGTAA...GTCATC 1461
      : : : : : | : : : : : | : : : : : | : : : : :
35 sPTyrSnaSnaGlnSerIleIleLysAlaGlyGluArgGlnIleGlyIle 51
1462 TTGATGATCAGGCGGACGATCAAGCAAAAGCAAGCCCTTTAGTAAT 1511
      : : : : : | : : : : : | : : : : : | : : : : :
52 His1leYse1SerSpsGlyAlaGlyValArgThAlaThrGlyThr 68
1512 CCGCTTGCTGTCAGCGGACGAGGCGGACGTCGAATGCGATTAATCACT 1561
      : : : : : | : : : : : | : : : : : | : : : : :
68 rIleLySvalSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn... 83
1562 TCACACCCGACAACTATTCGCGCTTCGCGCGGACGCTTGGAATTG 1611
      : : : : : | : : : : : | : : : : : | : : : : :
84 ....Pro1a1aG1uLeuArgPheGlnSnaGlySerValIThrSer 97
1612 AACGGCATTCGCTTCCTCCACCGCATCAAAATACCGATGAAGGCG 1661
      : : : : : | : : : : : | : : : : : | : : : : :
98 SerGly.....GlnLeuPheAsp1uGlyVala 106
1662 GATGATTGTCAACCAATCAAGACAAGAATCCACGTTACATTACAG 1711
      : : : : : | : : : : : | : : : : : | : : : : :
106 1.....ArgArGpHeuGlyThrValThrValIlysa 117
1712 GCATTA.....GATATTACTACACCGCGCATATACACACTTG 1752
      : : : : : | : : : : : | : : : : : | : : : : :
117 laGlyLySleuValAlaAspHisAlaThrLeuAlaSnValSerAspThr 133
1753 GATGCAAAAAAGAAATTGCC...TACACGTTGGTTGGCGAGAAAGA 1799
      : : : : : | : : : : : | : : : : : | : : : : :
134 ArgAspAspAspGlyIleAlaLeuTyValAla.....GlyIuGln.. 147
1800 TGCACCAAAACGACAGCGGCGCTCAATCTGAATTACCAACCGAAGAG 1849
      : : : : : | : : : : : | : : : : : | : : : : :
148 .....AlaGlnAlaSerIle 153
1850 CGGATGCGACTTACTGCTTCGCGGGAACAATTA.....AACGC 1893
      : : : : : | : : : : : | : : : : : | : : : : :
153 laAspSerThrLeuGlnGlyAlaGlyValArgValGluArgGlyAla 169
1894 AATTCACGCAAAACAAAGCTGTTTTCAGCGGACGACCGACAC 1943
      : : : : : | : : : : : | : : : : : | : : : : :
170 AsnValThrValGlnArgSerThrIleValAspGlyGlyLeu..... 183
1944 GCACGCTTACATCATTTAGGAAGCGGGTGCTCA...AAATGGAAGTA 1990

```

```
184 .....HistIeglyThrLeuGlnProLeuGlnProGlnAspL 196
1991 TCCACAGAGAGAAATCGTGGACAAACGATTGGATCGACCGC ..... 2034
196 euProProSerArgValValLeuGlyAspThrSerValThrAlaValPro 212
2035 .....ACATTAAAGCGGAAACCTCCA 2057
213 AlaserGlyAlaProAlaValValPheValPheGlyAlaAsnGlnLeuThr 229
2058 TMTTCAGGGCGGACAAAGCGGTTCGCCGCAATGTGCCAAAGTGAAG 2107
229 rValAspGlyGly ..... 233
2108 GCGATTGGCATTAAACAATCAGCCCAAGCAGTTTGGTGTCGACCG 2157
234 .....HistIethrGlyArgAlaAla ..... GlyValAla... 244
2158 CATCAAAAGCCACAAATCTGTACGTTGCGACTGGAGCGGTGTGACAAG 2207
244 ..... 244
2208 TTGTACCGAAAAACCATTAACGAGATTAAGTGTGCTTCATTGACCA 2257
245 .....AlaMetAspGlyAlaLeuAlaHisLeuGlnA 255
2258 AGACCGACATCAGAGCAATGTGACGCTTCCGATCAGCGTCATTAAAT 2307
255 rGAlaThrIleArg ..... 259
2308 CTCACAGGACTTGCACACTCAGCGCAATCTTAGTCAGCGGAGACAC 2357
260 .....ArgIAspAlaProAlaGlyGly ..... 267
2358 GCACATACGCTTAGCGCAAGCGCCAAACGCAACCTC ..... 2400
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282
2401 .....AGC 2403
282 LypheGlyProLeuLeuAspGlyTyrTyrGlyValAspValSerAspSer 298
2404 CTCGTGGCGAATGCCCAAGCAACATTTAATCAAGCCATTAAAGCGCA 2453
299 ThrValAspLeuAlaGlnSerIleValGlnAlaProGlnLeuGlyAlaAl 315
2454 CACATCGGCTTCGGCAATGCTTCAATTAACTAAGC ..... 2490
315 AlIeArgAlaGlyArgGlyAlaArgValThrValSerGlySerLeuS 332
2491 .....AACACGCGCTACAAACGCG ..... 2511
332 erAlaProHisGlyAsnValIleGlnThrGlyGlyAlaArgArgPhe 348
2512 .....AGTGTACGCTTCCGACAAAGCTAAAGC 2540
349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365
2541 AAGC .....GTAGCCGATT 2554
365 agIneGlyArgAlaLeuLeuTyrArgValLeuProGlnProValIlyLeuT 382
2555 CGCGACTCAAGCAATGTCCCTAGCGTAAGGCAAGTATTCATTATTT 2604
382 hrLeuAlaGlyGlyAlaGlnGlyGlnIAsp .....IleValAlaThr 396
2605 GAAAAACAGCCGCTTACCGGAAATACGCGGCGGCAAGATACGCAATT 2654
397 GluLeuProProIleProGlyAlaSerSerGlyProLeuAspValAlaLe 413
2655 ACACCTTAAAGACAGCAATGAGCGCTGCGGCGACGGAATTAGCA 2704
413 u...AlaSerGlnAlaArgTyrThr...GlyAlaThrArgAlaValAspS 428
2705 ATTAAACCTTGACAAAGCCCACTTACACTCAATCCGCTATCGACAC 2754
428 erLeuSerIleAspAlaThrThrValMetThrAsp ..... 440
2755 GATCGGCGAGCGCGCAACCGCGAGTGCAGGAGATCGCGCGCGCGC 2804
441 .....AsnSerAsnValGlyAlaLeuArgLeuAl 450
2805 TTGCGCGCGCTTCCCTATTATCCGTTACCGCCCACTTCGCGAAATCC 2854
450 aserAspGlySer...ValAspPheGlnGlnProAlaGlnAla...GlyA 465
2855 GTTTCACAGCGCTGACGTAACGCGCAATTAAGACGTCGAGGACATTC 2904
465 rGpHeuIlyValIleuMetValAsp...ThrLeuAlaGlySerGlyLeuPhe 480
2905 CGCTTATGTGCGAATCTTCGGCTACCGCAGCGGCAATTAAGCTGCG 2954
481 ArgMetAsnValPheAlaAspLeuGlyLeuSerAspIlyLeuValIle 497
2955 GGAAGATTCCGAAGGCACTTACACTTGGCTGTCAACATACCGGACAG 3004
497 lArgAspAlaSerGlyGlnHisArgLeuTyrValArgAsnSerGlySerG 514
3005 AACCCGTAAGTCTGACGCAATGACGCTAGTAGAGAGAAAGCAACACA 3054
514 lProAlaSerGlyAsnThrMetLeuLeuValGln .....Thr 526
3055 CCGCTGTCCGAAATCTTAATTCACCTGCAAAACGAA .....CAGCT 3098
527 ProArgGlySerAlaAlaThrThrPheThrLeuAlaAsnIlyAspIlyS 543
3099 CGATGCCGCGCATGGCGTTATCAGCTTATCCGAAA...GACGGGAGT 3145
543 lAspIleGlyTThrTyrArgTyrArgLeuAlaAlaAsnGlyAsnGlyI 560
3146 TCCGCTCGCATTAATCCGTCAAAGAACAGAGCTTCGCAAACTCGCG 3195
560 rPserLeu.Val ..... 563
3196 AAGCGGGAGAAACAGAGCGCGCTTGACGCAAAACAGGCAACACTTGC 3245
563 ..... 563
3246 CGCCAAACACAGCGGAGAAAGACAAACGCAACGCTTGACGCGCTGA 3295
563 ..... 563
3296 TTGCGCGCGGCGCAATGCCACCGAAAGCAGAAAGTGTGCGCAACG 3345
564 .....GlyAlaIlyAlaProProAlaProIlyAspProAlaPro...Gl 576
3346 GCCCGGAGGCGAGCGGAGAAATGCCGCAATTATGACGCGGAGAGAGA 3395
576 nProGlyProGlnProGlyProGlnProPro ..... 586
3396 GAAAAAAGCGGTGCAGCGGATTAAGACACCGCGCTTGCGGAAACAGCGG 3445
586 ..... 586
3446 AAGCGAAACCGCGCGGCTACACAGCGCTTCCCGCGCGCGCGCGCGC 3495
587 .....GlnProProGln.ProProGlnProProGlnP 597
3496 CGCGGAGATTGCGCAACCGCAGCCCAA .....CGCAACCCCA 3536
597 roPro.GlnProProGlnArgGlnProGlnAlaProAlaProGlnProPr 613
3537 ACCGCAAGCGGACACTGATCAGCGCTTATGCAATAGCGGTGAGATGAT 3586
613 oAlaGlyArgGlnLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrG 629
```



```
222 ..glyValIleSerValAsnIleGlySerIleSerLeuLeuAlaGlyIn 237
322 CAA.....TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGGCACAATAACGGCGCTATACAAATGTTGATTTGGTCGGAGGAA 385
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaIleGlyG 271
386 GCAAT.....CCCGATCAGACCGCTTTTCT 411
271 LysAsnIleAsnValArgAlaIleAlaThrIleArgAsnGlnGlyLeuSer 287
412 TACCAATTTGTGAAGAATAATTATTAAGCAGGAGCTACGGCCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGCGATTATCATATATGCGCGTTTGCACAAATTTGTCAAGATG 511
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTGAG.....ATGACCAAGTTATATGAT 543
303 IalysGluGlyGluAlaGluIleGlyValIleSerAlaGluAsnGln 319
544 GGGTGAATACGCTGATTTAATAATATACCGTATCGTGT.....CG 587
320 GluAlaIalysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGAGACA.....GGCAGACAAATAT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyIleThrLys 353
611 GCGCGTCTGATGAA...GACGACACCAATAC..... 639
353 euGlyGlyAspGluArgGlyGlyLysAsnGlyIleGluLeuAlaLys 369
640 .....CGGAAAGTTCATATCATATTTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLys 386
667 .....GCATATTCCTTGG.....CTCGTCGGTGGCATA 694
386 sGlyGlyPheAlaIleValIleArgGlyAspIleAlaLeuIleAspLysnI 403
695 CCTTGCACAAATGANTGAGTGTGGCAGCAGTCACTTAGTAGCGAA 744
403 IeAsnAlaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGTTTATTAACAAGAGAGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATCCCAAAAAGCAAAATG 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAAATGGGTATGCAAAACAGCAACCCCTATATAGCAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCAGCTAGTTCGTAAGATGTTCTATGATGAATCTTTGCTGG 944
436 .....LysGluThrLeuLeuAsp.....Phe 442
945 AGATACCATTCAGTATTTACGAA...CCACATCAAAATGGGAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGly 459
992 TTTTAAAGCAATATATAGCGCGCAAAAATGCATGCCAAACATATAA 1041
459 IeAsnAspGluPheProThrGlyThrGlyAlaIleSerAspProLysLys 475

1042 CACTATTCCTACCTTATAGATTAACACAGACCGTCAATTTGTT... 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnIle 492
1090 .....AATGTTTCTTTATCCGACAGCAGCAAGAGAAC 1120
492 uLysAsnAlaIlePheThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGCAGGTGGGCTCAACAGTTATCCACCCAGACTGAT 1170
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGGAAATAATTTCTTTATTTACAAAGGAAAGTGAATGTATCT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACACACAAATCAACCAAGCGCGCGGCTTTGATTTTGAGGGTAAT 1270
527 .....GluArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGCTTCGCCTAAAAACAAACGAGTGGCAGGCGCGGCTTCAT 1320
538 IeThrSerLysGlyGlyAsnLeuThrIleThrSerGlyGlyIlePheValAsp 554
1321 ATCAGTGAATGGCAGTACCGCTTACTTGAAAGTAAACGGCGTGCACACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CGGCTGTCCAAATAATCGCAAGGACGCGTGTGTTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnSer 573
1414 .....AAAGGGGAAACCAAGCGTGTGACGGTGGCGGACGTTAA 1455
573 AlaIaPheGluGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTGATCAGCAGCGGAGCATCAAGGCAAAAACAGCCTTAG 1505
590 IleVal..... 591
1506 TGAATGCGCTTGTCAGCGGAGGCGGACGTCGACGTAATGCCGATA 1555
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCGACAACTATTTGCGCTTCGCGCGGCGGACGTTG 1605
602 LysLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTTCCACCGCATTAATAATCCGATGA 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGCGGATGATTTGC.....AACCAATATACAGCAAAAGANT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACGTTACCATTCAGGCAATTAAGATATTAAGTACACCGGCATATAC 1743
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerLysThrPheIleThrSerHisAspSerHisThrAsnValSerAl 667
1747 ....AACTTGAT..... 1755
667 AleuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysThrIleSers 684
1756 .....AGCAAAAAGCAATTTGCC 1773
684 eAsnSerLysGlyLeuThrThrGlnIleTyArgSerSerAlaGlyValAsn 700
```



```

1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
3453 AACCCGGGGGCTACCAACGGCTTCCCGGGGCCGGCGCGCGGG 3502
1131 ..... 1131
3503 ATTTGCGGCAACCGACCCCAACCGCAACCCCAACCGACGCGAGCTG 3552
1132 ..... 1132
3553 ATGCGC...CGTATGCCATAGCGGTTTGAGTGAATTTTCCGCGAGCT 3599
1133 ..... 1133
1136 ThrSerLeuYsthrValAsnIleThrAlaSerGlyValThrThr..T 1152
3600 CAACAGCGTTTTCGCGTACAGACGAATTGGACCGGCTGTTTCCCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyAlaSerIleThrThr 1168
3650 ACCCGCGCAACGCGGTTTGGACAAAGCGGATCCGGGACACCAACTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerVal 1185
3700 CGTTCGC...AAGATTCCGGCGCTACCGCGCAACAAACCGACGCGCCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLys 1202
3747 AATCGGATGCGAGAAACCTCGCGAGCGCGGCGGCGATCCTGTTT 3796
1202 ergIgluAlaAsnValThrSerAlaThrGlyThrIleGly..... 1216
3797 CGCACACCGCGCGGAAACCTCGCGAGCGGCGGCGATCCTGCGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1232
3839 ..... 1232
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThr 1249
3861 TGCCGTTTTCGGCAAT.....ACGGCATCGGAGTTGCGACATCGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerIleThr 1265
3905 TCACGCGCGCGCGGTTTGTAGTACGCGGACGCTTTCAGACGCGATCA 3954
1266 SerThrLysGlyGlnValAspLeuIleAla.....GlnAsnGlySer 1280
3955 GGCAAAATCGCGCGCGGCTGCGATTCAGCGCATTCAGCGCAATACCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThr 1297
4005 CGCAGGTTTCGCGGATTCGCGATCGCAACCGCATCGCGCAACGCGCT 4054
1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313
4055 ATTTGCGTCAAA 4066
1314 IleAsnAlaLys 1317

```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4

```

seq_documentation_block:
; Sequence 4, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mathare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington

```

```

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

```

```

alignment_scores:
Quality: 274.50 Length: 1488
Ratio: 0.419 Gaps: 73
Percent Similarity: 44.019 Percent Identity: 19.220

```

alignment_block:
US-09-303-518d-653 x US-08-302-832-4 ..

```

Align seg 1/1 to: US-08-302-832-4 from: 1 to: 1477
178 AATAAGGAGAGTTTTCAGTTCGCGGCGGAAAGATATGAGTTTACACAA 227
199 AsnIleGlyLeuIleThrValGly...LysAspGlySerVal..... 211
228 AAAAGGAGAGTTTTCAGTTCGCGGCGGAAAGATATGAGTTTACAT 277
212 .....AsnLeuIleGlyGlyValLysValLysAsnGlu..... 221
278 TTTCTGTGATTCGCGTACGCG.....GTGCGGCGATTGCGGCGCAT 321
222 ..GlyValIleSerValAsnGlyLysSerIleSerLeuLeuAlaGlyGln 237
322 CAA.....TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProGlnIleThrTyrSerIleAl 254
336 CGTGCACATTAACGCGCGCTATTAACATGTTGTTGTCGCGAGGCAA 385
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAT.....CCGATTCAGCGCGCTTTCT 411
271 LysAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATTTGGAAGAATAATTTAAGCAGGAGTAAGCGCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGCGCGGATTAACATATATGCGCGGTTTGCACAAATTTGTCAGATG 511

```



```

1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
      ::::: ||| ::::: ||| :::::
3700 CGTTCCG...AGATTCCGCGCTACCGCCAAACACACCTCGGCCA 3746
      ::::: ::::: ::::: ::::: :::::
1185 rAlaThrValAspLeuThrThrIleSerGlySerIleGluAlaLys 1202
      ::::: ::::: ::::: ::::: :::::
3747 AATGGGATGACAGAAACCTCGACGCGCGGCGGCGCATCCTGTTT 3796
      ::::: ::::: ||||| ::::: :::::
1202 eTgIgluAlaAsnValThrSerAlaThrGlyThrIleGly 1216
      ::::: ::::: ::::: :::::
3797 CGCAACAACCGACCGCAACACCTTCGACGACGACATCGCA..... 3838
      ||| ::::: ||| ::::: ||| :::::
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnIleGlyAsp 1232
      ::::: ::::: ::::: :::::
3839 .....ACTCGGACGCGCTTGGCCACGG 3860
      ::::: :::::
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGlyAlaAlaThr 1249
      ::::: ||| :::::
3861 TGCCGTTTGGGCAAT.....ACGGCATCGGAGGTTGCACATCGGCA 3904
      ::::: ||||| ||| ::::: ||||| :::::
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerIleThr 1265
      ::::: ||||| ::::: ||| ::::: ||||| :::::
3905 TCAGCGCGGCGCGCGGCTTTAGTACGCGACCTTCGACGACGACATCAGA 3954
      ||| ::::: ||| ::::: ||| ::::: ||| :::::
1266 SerThrIleGlyGluValAspLeuLeuAla.....GlnAsnGlySer 1280
      ::::: ||| ::::: ||| ::::: ||| :::::
3955 GGCAAATCCGCGCGCGCTGCTGCATTACGCGCATTCAGCAAGAACG 4004
      ||| ||| ||| ::::: ||| ::::: |||
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThr 1297
      ::::: ||| ::::: ||| ::::: ||| :::::
4005 CGCAGGTTTCGGGATTCGCGCATCGCAACCGCACATCGCGCAACGGCT 4054
      ::::: ||| ||| ::::: ||||| :::::
1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313
      ::::: ||| ||| ::::: ||||| :::::
4055 ATTTGCTCAAA 4066
      ||| ::::: |||
1314 IleAsnAlaLys 1317
      ::::: |||
seq_name: /cgn2_6/ptodata/1/1aa/5b_COMB.pep:US-08-530-198-4
seq_documentation_block:
; Sequence 4, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810

```

```

; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-4

alignment_scores:
      Quality: 274.50      Length: 1488
      Ratio: 0.419      Gaps: 73
      Percent Similarity: 44.019      Percent Identity: 19.220

alignment_block:
US-09-303-518D-653 x US-08-530-198-4 ..
Align seg 1/1 to: US-08-530-198-4 from: 1 to: 1477

178 AATAAGCAAGCAAGTTTGCAGTCGGCGCGCAAGATATTGAGTTTACACAA 227
      ||| ||| ::::: ||||| |||||
199 AsnHISGlyLeuIleThrValGly...LysAspGlySerVal..... 211
      ::::: ||||| :::::
228 AAAAGGGAGCTTGTCGCAATCGATGACGAAACCGCCGATGATTGATT 277
      ::::: ||||| :::::
212 .....AsnLeuIleGlyGlyValIleLysAsnGlu..... 221
      ::::: ||||| :::::
278 TTTCTGTGTATCGGCTAACGCG.....GTGCGCGCATTCGGCGCGCAT 321
      ::::: ||||| ||||| ::::: ||||| :::::
222 ...GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGly 237
      ::::: ||||| :::::
322 CAA.....TATATGTGAC 335
      ::::: ||| :::::
238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254
      ::::: ||| :::::
336 CGTGCACATTAACGGCGCTATACATGTTGTTGCGGAGGAA 385
      ::::: ||||| ::::: ||| ::::: ||||| :::::
254 AlaIleProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
      ::::: ||||| :::::
386 GCAAT.....CCGATCAGCACCGCTTTCT 411
      ::::: ||| :::::
271 LysAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287
      ::::: ||||| :::::
412 TACCAATTTGCAAAAGAAATATATATAAGCAGACGACGACGACATCC 461
      ::::: ||||| :::::
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
      ::::: ||||| :::::
462 TTATGCGGCGATTTATCATATCCCGGTTTGCACAAAATTTGCACAGATG 511
      ::::: ||||| :::::
298 .....AsnIleValIleSerA 303
      ::::: ||| :::::
512 CAGAACCTGTGAG.....ATGACAGATTATATGAT 543
      ||| ::::: ||| :::::
303 LysGluGluGluAlaGluIleGlyValIleSerAlaGlnAsnGln 319
      ::::: ||||| :::::
544 GGGTGAATACGCTGATTAAATACCTGATCGTGT.....CG 587
      ||| ::::: ||| ::::: ||||| :::::
320 GlnAlaLysGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
      ::::: ||| ::::: ||| :::::
588 AATCGGAGA.....GGCAGACAAATATT 610
      ::::: ||||| :::::
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
      ::::: ||||| :::::
611 GCGGCTGTATGAA...GACGAACCCAAATAC..... 639
      ::::: ||||| ||| ::::: |||||
353 euGlyGlyAspGluArgGlyGlyLysAsnGlyIleGlnLeuAlaLys 369
      ::::: ||||| :::::
640 .....CGCGAAGTTTCATTCATTTGCAAG..... 666
      ::::: ||||| :::::
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLys 386
      ::::: ||||| :::::

```

```

667 .....GCATATCTGTG.....CTGCTGGTGCATA 694
668 ||||| ||||| ||||| |||||
386 sGlyGlyPheAlaIleValIleThrGlyAspIleAlaIleLeuIleAspIleGlySni 403
695 CCTTGGCAAAATGATCAGTGGTGGCAGTCACTAGTAGTACGAA 744
696 ||||| ||||| ||||| |||||
403 IeAsnAlaIleGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGTTTTTTTACCAACAGAGAGCTCATTTGG 794
746 .....AspIleAlaIleThrGlyGlyPheValIle 419
795 CGACAGTGGCTCCCAATGTTTATCTATGATCCCAAAAGCAAAAGTGT 844
796 ||||| ||||| ||||| |||||
419 uThrSerGlyHisAspLeuPheIleIleGlyAsp..... 429
845 TAATTATGGGGTATTGGCAACAGGCAACCCCTATATAGCAAAAGCAAT 894
846 ||||| ||||| ||||| |||||
430 .....AsnAlaIleValAspAla..... 435
895 GGCCTTCAGTAGTTCGTAAAGATTGTTCTATGATGAATCTTTGCTGG 944
896 ||||| ||||| ||||| |||||
436 .....LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATTTCTACGAA...CCACATCAAAATGGGAATACT 991
946 ||||| ||||| ||||| |||||
442 eAspAsnValSerIleAsnAlaIleGluAspProLeuPheAsnAsnThrGly 459
992 TTTTAAACGACATATATATAGCGGCAAAATCGATGCCAAACATATAA 1041
993 ..... 459
459 IeAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLysLys 475
1042 CACTATTCCTACCTTATAGATTAAAAACAGAACCTTCATTTGTTT.. 1089
1043 ..... 475
476 AsnSerGlyLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyLe 492
1090 .....AATGTTCTTATCCGACACAGCAAGACAGAAC 1120
1091 ..... 492
492 uLysAsnAlaIleThrPheMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGACAGTGGGCTGACAGTTATCGACCCAGACGTGAT 1170
1122 .....ValAsnSer.....SerIleAsn 513
508 ..... 513
1171 AATGGAGAAATATTTCTTATGACAAAGAAAGTGAATGATACT 1220
1172 ||||| ||||| ||||| |||||
514 IleGlySerAsnSerIleLeuIleLeuHisSerGly..... 526
1221 TACCAAGACATCTCAACCAAGCGCGGCTTTGATTTGAGGGTAAAT 1270
1222 .....GlnArgGlyGlyValGlnIleAspIleAspI 538
527 ..... 538
1271 TTACGGTCTGCGCTAAACACAGCAAGCGTGGAGCGGGCGGTTCAT 1320
1321 ..... 538
538 IeThrSerIleGlyGlyAsnLeuThrIleThrSerGlyGlyIleThrValAsp 554
1321 ATCAGTATGCGCAGTACCGCTTACTTGGAAATGAACGCGTGGCAACGA 1370
1322 ..... 554
555 ValHisLysAsnIleThr..... 560
1371 CCGCGTCTCCAAATCGCAAAAGCAGCTGCTGTTCAAGCC..... 1413
1372 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsp 573
561 ..... 573
1414 .....AAAGGGAAACCAACAGCTCGGTGACGCGTGGCGAGCGGTA 1455
1415 ..... 1455
573 AlaIlePheGlnGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLys 589
1456 GTCATCTTAATATGACGAGCGGAGCATCAAGCAAAACAGCTTTAG 1505
1457 ..... 1505
590 IleVal..... 591
1506 TGAATATGCGCTGGTCAAGCGGAGGCGGAGCTGTCACATGATGCCGATA 1555

```

```

592 .....AlaIleGlyThrValThrIleThrGlyGln 602
1556 ATCAGTTCAACCCCGACAAACCTATTTGCGGCTTGGCGGAGCTTG 1605
1557 ..... 1605
602 Lys.....AspPheArgAlaAsnAsnVal 610
1606 GATTGAACGGGCATTCGCTTTCGTTCCACCGCATTTCAAAATACGATGA 1655
1607 ..... 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGCGCGATGATGTC.....AACCAATCAAGCAAAAGAT 1693
1657 ..... 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSer 634
1694 CCACGTTACCATTTACAGCAATTAAGATATTATTCACACCGGCAATAC 1743
1695 ..... 1743
634 LThrIleAsnIleSerGlyAsnIleThrIleAsnIleThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyrThrPheIleThrSerHisAspSerHisThrAsnValSerAl 667
1747 .....AACTGGAT..... 1755
667 AleuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrIleSer 684
1756 .....AGCAAAAGAAATTTGCC 1773
684 eAsnSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyValAsn 700
1774 TACACGGTTGGTTGGCGAGAAAGATGCACCAAAACAGACGCGCGCT 1823
1775 .....ValAsnGlyAsnMe 708
701 PheAsnGly..... 708
1824 CAATCTGAATTTACCAACCGGAAGCGGATCGCACTTACTGCTTTCCG 1873
1825 ..... 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAATTTAAC.....GGC 1893
724 roAsnIleAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCCAGCAACAA...AACGCAAACTGTTTTCAGCGGCGACCGAC 1940
1895 ..... 1940
741 AsnIleThrAlaIleThrGlyGlySerValPhePheAsp..... 753
1941 ACCGACAGCTAC...ATCATTTAGGAAGCGGCTGTCAAAATGGAAG 1987
1942 ..... 1987
754 .....IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAACAGAGAAATCGTGGGACACAGATTGATGACCGGACA 2037
1989 ..... 2037
764 IuLeuLysMetSerGluIleAsnIleSerAsp..... 774
2038 TTTAAAGCGAAACTTC.....CATATTCAGGCGGACGAC 2075
2039 ..... 2075
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGTTTCCCGCATGTTGCCAAAGTGAAGCGCATTTGCAATTAACA 2125
2077 ..... 2125
789 abPe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCACAGCATTTTCGCTGTCGACCGCATCAACGACACATTC 2175
2127 ..... 2175
799 ..... 799
2176 TGTACAGTTCGACTGACGCGGTCTGACAACTTACGAAAAAACCAT 2225
2177 ..... 2225
800 AlaThrAsnSerAsnPe..... 805
2226 TACCGACATTAAGTATGTTGTTCAATTGACCAAGACGACATCAGAGCA 2275
2227 ..... 2275

```

806SerLeuArgInThrLysAspPheT 815
2276 ATGACGCTTGGCGATCAGCTCATTTAAATCTCACAGGACTTGGACA 2225
815 yraspglytYrAlaArgAsnAlaIleAsnSerThrTyraSnIleSerIle 831
2326 CTCACGCGCAATCTTATGCGAGGGGAGACAGCACTATACGTTACGCG 2375
832 LeuGlyLysValThrLeuGlyGly..... 840
2376 CAACGCCAACCAACGCGCACTCAGCTCGGGCAATGCCAACAA 2425
841GlnAsnSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTAAATCAAGCC.....ACATTAAACGGCAACACATGGCTTCG 2466
852 hrIleGlyLysAlaAlaAsnValThrLeuGlnAlaAsnAlaProAsn 868
2467 GACAATGCTTCATTAACTTAAGCAACAGCCGTACAAACGCGACTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerIle 882
2517 GACGCTTCCGACACGCTAAGGCAAAACGTAAGCCATTCCGACTCAAG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyLysAlaAsnAlaAspIleLysG 899
2567 GCAATGTCTCCCTAGCCGATAGGCACTATTCAT.....TTTGAA 2607
899 LysnLeuThrIleSerGlnSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCGGCTTACCGGCAAAATACAGCGGGGAGATACGGCATTTACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnGlyThrAlaGlnIleAs 932
2658 CTTAAAGACAGCAATGACGCTGCGGTGGGACGGAATTAGCGATT 2707
932 nIleThrGlnGlyValIleLysLeuGlyAsnValThrAsnAspIleLys 949
2708 TAAACCTTGACAAACGCCACCTTACACTCAATTCGCCCTATCGACAGAT 2757
949 euAsn.....IleThrThrAla..... 955
2758 GCGGAGCGCGCAACCGGCAAGTGGCGAGATGCGCGCGCGCTTC 2807
956LysArgAs 958
2808 GCGCGCTTCCCTATATTCGTTACGCCGCCAATTGCGCAAAATCCGCTT 2857
958 nGlnArgSerIle..... 962
2858 TCAACACGCTGACGTAACGCAATTTGAACGCTCAGGAGACATTCGCG 2907
963IleGlyLysPheIleAsn..... 969
2908 TTTATGTCGAACCTTTCGGCTACCGGAGCGGCAATTTGAAGTCGGCGA 2957
970LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTC.....G 2965
978 pSerAsnAsnAspAlaGlnIleGlnIleGlyLysAsnIleSerGlnLysG 995
2966 AAGCAGCTTACACCTTGCT.....GTCAACATTAACCGGCAACGAA 3006
995 LngLysnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCCGTAAGTCTGAGCAATGACGTA.....GTGGAAGGAAAGA 3047
1009LysGlnIleThrIleLysLysGlyLysAspGlyLys 1021
3048 C..AACACACGCTGTCCGAAATTTAATTTCACCTTCGAAACGAAAC 3094
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys..... 1035

3095 ACGTGATGCGCGCGATGCGCTTATCACTTATCCGCAAGACGCGGAG 3144
1035 1035
3145 TTCGCGCTGCATTAATCCGGTCAAGAACAGAGAGCTTTCGCAAACTCGG 3194
1036ThrLysGlnLeuLysLeuThrGlnAspLeuSe 1046
3195 CAAGCGGAGAAACAGAGCGCCCTTACAGCGCAAAACAGGACACACTTG 3244
1046 rIleSerGlyPheAsnLysAlaGlnIleThrAlaLysAspGlyArgAspL 1063
3245 CCGCCAAACACAGCGGGGAAAAAGAACACGCGCAAGC..... 3282
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGlnAlaLysThr 1079
3283CTTGACGCGCTGATTGGCGCC...GGCGGCA 3311
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGCACCGCAAAAGCGAAGAGTGTCCGACACCGCGCGCGACGACGG 3361
1096 nValThr.....LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111
3362 GGGAAATGCGCGCATTTATGACGGCGGAGAGCAAGAAAAACGGGTGCGG 3411
1111 LysArgIleSer.....Asn 1115
3412 GCGGATTAAGACACCGCCTTG.....GCGAAACACCGGAAAGCGA 3452
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGlyVal.. 1131
3453 AACCGCGCGGCTACACCGCCTTCCCGCGCGCGCGCGCGCGCGG 3502
1131 1131
3503 ATTTGCCGACACGCGACGCCCAACCGCAACCCCAACCGGAGCGGACCTG 3552
1132AsnLysAspIle 1135
3553 ATACAG...CGTTATGCCAATAGCGTTTGAAGTAATTTCCGACAGCT 3599
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGlnLysValThrThr..T 1152
3600 CAACGCGTTTTCGCCGTACAGAGCAATTTGACCGCGCTGTTCGCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCGCGCAACGCGCTTGGACAAAGCGCATCCGGGACACCAACACTTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTGCG...AGATTTCGCGCCTTACGCGCAACAAACGACCTGCGCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGlnAlaLys 1202
3747 AATCGGTATGCAAGAAACCTCGGACGCGCGCGCTGCGCATCTGTTT 3796
1202 erGlyGlnAlaAsnValThrSerAlaThrGlyThrIleGlyLys 1216
3797 CCGCAACACGCGGAAACACCTTCGACGACGCGATCGGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839ACTGCGACGCGCTTGGCCGCG 3860
1232 uThrValGlyAsnGlyAlaGlnIleAsnAlaThrGlnGlyAlaAlaThrL 1249
3861 TGCCTTTTCGGCAAT.....ACGGCATCGGAGGTTTCGACATCGGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGlnAlaGlySerSerIleThr 1265


```
845 TAATTATGGGCTATTGCAACAGCGCAACCCCTATATAGAAAAAGCAAT 894
      |||:::|||||:::
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCCAGTAGTTCGTAAGATTGGTTCATGATGAATCTTGGTCG 944
      |||:::|||||:::
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATTTCTACGAA...CCACATCAAAATGGCAAAATACT 991
      |||:::|||||:::
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459
992 TTTTAAACGACAAATAATATGCGCCAGAGAAAATCGATGCCAAACATATA 1041
      |||:::|||||:::
459 leAsnAspGluPheProThrGlyThrGlyAlaIleAsnAspProGlyLys 475
1042 CACTATTCTCTACCTTATAGATTAAACACGACCGCTCAATGTGT.. 1089
      |||:::|||||:::
476 AsnSerGluLeuLysThrThrIleuThrAsnThrThrIleSerAsnTyLe 492
1090 .....AATGTTCTTTATCCGACAGACAGACAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGACAGGTGGGTCAACAGTTATCCACCCAGCTGAAT 1170
      |||:::|||||:::
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAAATATTCTTATATGCAAGAAAGTAAGTGAATTGATACT 1220
      |||:::|||||:::
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACCAACACATCAACCAAGCGCGCGGCTTGTATTTAGAGGTAAT 1270
      |||:::|||||:::
527 .....GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGCTCGCCTAAACACAGCAAAACGTCGACGCGCGCGCTTCAT 1320
      |||:::|||||:::
538 leThrSerLysGlyGlyAsnLeuThrIleLeuSerGlyGlyTrpValAsp 554
1321 ATCAGTATGGCAGTACCGTTACTTGAAGTAACGCGCTGGCAACGA 1370
      |||:::|||||:::
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCAAAATCGCAAAAGGACGCTGCTGTTCAAGCC..... 1413
      |||:::|||||:::
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnSerV 573
1414 .....AAAGGGAACCAACGAGCTCGGTCAAGCTGGGCGACGCTAAA 1455
      |||:::|||||:::
573 alaIaPheGlnGlyGlyAsnAsnLysalaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTAGTCAGCAGGGGAGCATCAAGCAAAACAAAGCCTTTAG 1505
      |||:::|||||:::
590 IleVal..... 591
1506 TGAATCGCGCTGGTGCACGCGAGGAGGAGCTGCAACGTAGTCCGATA 1555
      |||:::|||||:::
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCGACAAACTATTTTCGCTTTCGCGGAGGACGTTTG 1605
      |||:::|||||:::
602 LysLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCTTGGCTTTCCTTCACCCGATTCAAATACCGATGA 1655
      |||:::|||||:::
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGGCGATGATGTC.....AACCAATCAAGACAAAGAAAT 1693
      |||:::|||||:::
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTTACAGCAATAAGATATTACTACACCGGCAATTAAC 1743

634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
      |||:::|||||:::
1744 AAC..... 1746
651 AsnThrSerTyTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 .....AACTGGAT..... 1755
667 alaLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyIleSerS 684
1756 .....ACGAAAAAAGAAATGGCC 1773
684 eAsnSerLysGlyLeuThrThrGlnTyArgSerSerAlaGlyValAsn 700
1774 TACACGCTTGTTGGTGGCGAAGATGCAACCAACCAAGACAGGCGGCT 1823
      |||:::|||||:::
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTTACACACCGCAAGAAAGCGGATCGACTTACTGCTTCG 1873
      |||:::|||||:::
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAAATTTAAAC.....GGC 1893
724 roAsnGluAsnMeAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATACCGCAACA...AACGGCAACCTGTTTTCAGCGGACAGCCGAC 1940
      |||:::|||||:::
741 AsnIleThrAlaThrGlyGlySerValPhePheAsp..... 753
1941 ACCGACAGCCTAC...AATCATTTAGAACGCGGCTGCAAAAATGGAAG 1987
      |||:::|||||:::
754 .....IleTyraIaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAACAGAGAAATCGTGGGACACACGATTTGATCGACGCGACA 2037
      |||:::|||||:::
764 IuLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGGAACCTTC.....CATATTCAGGCGGACCAAGC 2075
      |||:::|||||:::
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGTTTCCCGCAATGTTGCCAAAGTGAAGCGCATTTGCAATTAAGA 2125
      |||:::|||||:::
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCAGACAGTTTTCGGTGTGCGACCGCATCAAAAGCCACACAATC 2175
      |||:::|||||:::
799 ..... 799
2176 TGTACACGCTTGACGCTGACGCGGTCTGCACAGTTTACCGAAAAACAT 2225
      |||:::|||||:::
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACGATTAAGTGATTTGCTTCATTGAGCAAGACCGACATCAGAGCA 2275
      |||:::|||||:::
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCCTTCCCGATCAGCCTCATTTAAATCTCACAGACTTCCACA 2325
      |||:::|||||:::
815 yTrAspGlyTyraIaArgAsnAlaIleAsnSerThrTyraAsnIleSerIe 831
2326 CTCAAAGCAATCTTAGTCAGCGGAGACAGACGACTATACGGTTACGCG 2375
      |||:::|||||:::
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAACGCAACCCAAACGCAACCTCAGCCTGCTGGGCAATGCCCAAGCA 2425
      |||:::|||||:::
841 .....GlnAsnSerSerSerIleThrLysAsn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAACGCGGACACATCGGCTTCG 2466
      |||:::|||||:::
```

852 hrIleGluYsAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsn 868
2467 GACAATGCTTATTATTCTAAGACAACAGCCCTACAAAAGCCACTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleYsLeuGlySerLe 882
2517 GACCTTTCGACAAAGCTAAGGCAACGTAAAGCCATTCCGCACTCAAG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGluAlaAsnAlaSpIleYsG 899
2567 GCATGTCTCCCTAGCCGATAGGCAGTATTCAT.....TTTGAA 2607
899 LysnLeuThrIleSerGluSerAlaThrPheYsGlyYsThrArgAsp 915
2608 AACGCGCCCTTACCGGAAAAATACGGCGGCAAGGATACGGCATACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAGACAGCAATGAGCTGCCGTGCGGCAAGAAATTAGCAATT 2707
932 nIleThrGlnGlyValIleYsLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAAAGCCACCTTACACTCAATCCGCTATCGACAGAT 2757
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGCAGGCGGCAACAGCGGAGTGGGAGATGCGCGCGCGCTTC 2807
956LysArgAs 958
2808 GCGCGCTTCCTATTATTCCTTACGCGCAACTTCGGCAAGATCCGCTT 2857
958 nGlnArgSerIle..... 962
2858 TCAACAGCTGACGGTAAAGGCAATTGACGCTCAGGAGACATTCCGC 2907
963IleGlyYsPheIleAsn..... 969
2908 TTTATGTCGAACCTTCGCTACCGGCAAGTGAAGTGGCGGA 2957
970LysYsGlySerLeuAsnIleThrAs 978
2958 AAGTTC.....G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyYsAsnIleSerGlnYsG 995
2966 AAGGACTTACACTTGCT.....GTCAACAATTACCGGCAAGCAA 3006
995 uGlyYsAsnLeuThrIleSerSerAspYsIleAsnIleThr..... 1008
3007 CCCGTAAGCTCGAGCAATTGACGTA.....GTGAAAGGAAAAGA 3047
1009LysGlnIleThrIleYsYsGlyIleAspGlyGluAs 1021
3048 C...AACACACCGCTGTCCGAAATCTTAATTACCCCTGCAAAACGAA 3094
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleYs..... 1035
3095 ACGTCGATGCGCGGCGATGCGTATATCCGCAAGACGGCGAG 3144
1035 1035
3145 TTCGCGCTGCAATAATCCGCTCAAGAACAGAGCTTCGCAACACTCG 3194
1036ThrYsGluLeuYsLeuThrGluAspLeuSe 1046
3195 CAAGCGGGAACAGAGCGCGCTTGACGCGCAAAACAGGCAACACTTG 3244
1046 rIleSerGlyPheAsnYsAlaGluIleThrAlaYsAspGlyArgAspL 1063
3245 CCGCCAAACAGACAGGCGGAAAAAGCAACGCGCAAGC..... 3282
1063 euThrIleGlyYsAsnSerAsnAspGlyYsAsnSerGlyAlaGluAlaYsThr 1079
3283CTGACGCGCTGATTGCGCC...GGCGCAA 3311
1080 ValThrPheAsnAsnValYsAspSerYsIleSerAlaAspGlyHisAs 1096
3312 TGGCACGGAAGAGCAAGATGTGCGCAACCGCGCGGCAAGCGCG 3361
1096 nValThr.....LeuAsnSerYsValYsThrSerSerSerAsnGlyG 1111
3362 GCGAAATCCCGCATTAATGACGGCGGAGAGAGAAAAACGGGTGCG 3411
1111 LysArgGluSer.....Asn 1115
3412 GCGGATTAAGACACGCGCTTG.....GCGAAACAGCGGAAAGCGGA 3452
1116 SerAspAsnAspPheThrGlyLeuThrIleThrAlaYsAsnValGluVal.. 1131
3453 AACCGCGCGGCTACCGCCTTCCCGCGCGCGCGCGCGCGG 3502
1131 1131
3503 ATTGCGGCAACGCGAGCCCAACCGCAACCCCAACGCGAGCGAGCTG 3552
1132AsnYsAspIle 1135
3553 ATCAGC...CGTTATGCCAATAGCGGTTTGATGATTTTCGCGACGCT 3599
1136 ThrSerLeuYsThrValAsnIleThrAlaSerGluYsValThrThr.T 1152
3600 CAACAGCGTTTCGCGGTACAGAGCAATTGACCGCGCTTGGCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyYsAlaSerIleThrThr 1168
3650 ACCGCGCAACGCGCTTGACAAAGCGGATCCGGACACCAACACTAC 3699
1169 YsThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTGCG...AAGATTCCGCGCTTACCGCAACAAACCGACCTCGCCA 3746
1185 rAlaThrValAspLeuThrThrYsSerGlySerYsIleGluAlaYsS 1202
3747 AATCGGTATGCAAGAAACCTCGCGCGCGCGCGCTGCGCATCGTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyYs..... 1216
3797 CGCACAACGAGACCGGAACACCTTCGACGAGCATCGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839ACTGCGCACGCGCTTGCCACGG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGlnGluYsAlaAlaThrL 1249
3861 TGCGCTTTCGGGCANT.....ACGGCATCGGAGGTTCGACATCGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265
3905 TCAGCGCGGCGCGGTTTGTAGTACGCGCAAGCTTCACAGCGCATACA 3954
1266 SerThrYsGlyGluValAspLeuAla.....GlnAsnGlySerI 1280
3955 GGCAAAATCCGCGCGCGCTGCTGCTTACGGCATTCAGGCAAGATACG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CCGAGGTTTCGGGATTCGGCATCGAACCGCAATCGCGGCAAGCGCT 4054
1297 euThrThrValAlaGlySerAspIleYsAlaThrSerGlyThrLeuVal 1313
4055 ATTTCGTCCAAA 4066
1314 IleAsnAlaYs 1317

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-728-470-4

seq_documentation_block:

; Sequence 4, Application US/08728470

; Patent No. 5928651

; GENERAL INFORMATION:

; APPLICANT: Batenkamp, Stephen J

; TITLE OF INVENTION: High Molecular Weight Surface Proteins

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Shoemaker and Matlare, Ltd.

; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

; STREET: Bldg. 1

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202-0286

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/728,470

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/302,832

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/US93/02166

; FILING DATE: 16-MAR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9205704.1

; FILING DATE: 16-MAR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Bektresser, Jerry W

; REGISTRATION NUMBER: 22,651

; REFERENCE/DOCKET NUMBER: 1038-633

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 415-0810

; TELEFAX: (703) 415-0813

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1477 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-728-470-4

alignment_scores:

Quality: 274.50 Length: 1488

Ratio: 0.419 Gaps: 73

Percent Similarity: 44.019 Percent Identity: 19.220

alignment_block:

US-09-303-518D-653 x US-08-728-470-4 ..

Align seg 1/1 to: US-08-728-470-4 from: 1 to: 1477

```

178 AATAAGGCAAGTTTGCAGTCGGGCGGAAGATATTGAGTTTACAAACA 227
    ||| ||| :|||:||||| ||||| :|||
199 AsnIleGlyLeuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAGGAGGTGGTGGCGCAATCGATGACGAAGCCCGATGATTGATT 277
    :|||:||||| :|||: |||
212 .....AsnLeuIleGlyLysValLysAsnGlu..... 221
278 TTTTGTGGATGCGGTACGCGG.....GTGGCGGATTGGCGGCGAT 321
    :|||:||||| ||||| :|||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237

```

```

322 CAA.....TATATTGTGAG 335
    :|||
228 LysIleThrIleSerAspIleIleAsnProThrIleThrIleThrIleAl 254
    :|||
336 CGTGGCACATACAGCGCGCTTAAACATGTTGATTTGGTCGAGGAGAA 385
    :|||:||||| :|||
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
366 GCAT.....CCGATTCAGACCGGTTTCT 411
    :|||
271 LysNIIeAsnValArgAlaAlaThrIleatgAsnGlnGlyLysLeuSer 287
412 TACCAATTGTAAAGAAATATTAATTAAGCAGGACTAAGCGCATCC 461
    :|||:||||| :|||
288 AlaAspSerValSerLysAsp.....LysSerIle..... 297
462 TTATGCGCGGATTAATCATATGCGCGCTTTCACAAATTTGTACAGATG 511
    :|||:||||| :|||
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTGAG.....ATGACCGATTATATGAT 543
    :|||:||||| :|||
303 IalysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319
544 GGGTGGAAATACGCTGATTTAATTAATACCGCTGCTGT.....CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValIleLeuLys 336
588 AATCGAGCA.....GGCAGACAAATAT 610
    :|||:||||| :|||
336 sThrGluValAlaValIleAspLeuSerGlyLysGluGlyGluThrIle 353
611 GCGCGTCTGATGAA..GACCAACCAATATC..... 639
    :|||:||||| :|||
333 euGlyGlyAspGluArgGlyGlyGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGGAAAGTTCAATCATATTCACAGC..... 666
    :|||:||||| :|||
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGly 386
386 sGlyGlyPheAlaIleValIleArgLysAspIleAlaLeuIleAspLys 403
695 CCTTTCACAAATGATGATCAGTGTGGCACATCACTTAAGTAGCGAA 744
    :|||:||||| :|||
403 IeAsnAlaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGCTTTTTCACACAGAGGCTCATTTGG 794
    :|||:||||| :|||
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGTCCACCATGTTATCTATGATGCCAAAGCAAAAGTGT 844
    :|||:||||| :|||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGGTATTCGCAACAGCAACCCCTATATAGCAAAAGCAAT 894
    :|||:||||| :|||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTCCAGCTAGTTGCTAAGATGTTGTTATGATGATAATCTTGCTGG 944
    :|||:||||| :|||
436 .....LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATTTCTACGAA...CCACATCAAAATAGGCAAACT 991
    :|||:||||| :|||
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnThrGly 459
992 TTTTAAACGACAATATATATGCGCAGCAAAATCGATGCCAAACATAA 1041
    :|||:||||| :|||
459 IeAsnAspGluPheProThrIleThrGlyAlaIleSerAspProLysLys 475

```

1042 CACTATTCTACCTTATAGATTAAAAACAGACCGTCAATTGTT.. 1089
: : : : :
476 AsnSerGluLeuLysThrThrLeuAsnThrThrIleSerAsnTyrLe 492
1090AATGTTTCTTATCCGAGACAGCAGACAAAC 1120
492 ULysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGTGCAGGTGGGGCAACAGTTATCCACCACAGTAAT 1170
508ValAsnSer.....SerIleAsn 513
1171 AATGGAGAAATATATTCCTTATGACAAAGAAAGCTGAATGTACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACGACGACATCAACCAAGCGGGGGGTTGTTATTTGAGGGTAAAT 1270
527GlnArgLysLysLysValGlnIleAspLysAspI 538
1271 TTACGCTCTCGCTAAACAAACAGACGCGGCGGCGGCTTCAT 1320
538 IetHisSerLysGlyLysAsnLeuThrIleTyrSerGlyLysTrpValAsp 554
1321 ATCAGTGATGGCAGTACCGCTTACTTGAAAGTAAACGCGCTGCCAAACA 1370
555 ValHisLysAsnIleThr..... 560
1371 CCGGCTGCCAAATCGGCAAGGCGCTGCTGCTCAAGCC..... 1413
561LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnVal 573
1414AAGGGGAAACCAACAGCTCGCTCAAGCTGGGCGAGGCTTAA 1455
573 AlaIlePheGluGlyLysAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTAGATCAGACGGCGGACATCAAGCAAAACAGCCTTTAG 1505
590 IleVal..... 591
1506 TGAATCGCTTGTGTCAGCGGCGGCGGCTGCACTGATGATCCGATA 1555
592AlaGlnGlyThrValThrIleThrLysGluG 602
1556 ATCAGTTCACCCGACAAACTATTTCCGCTTCCGCGGCGAGCTTTG 1605
602 Lys.....AspPheArgAlaAsnVal 610
1606 GATTTCAGCGGCGATTCGCTTCCGTCACCGCATTCAAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGCGCATGTTGTC.....AACCAATCAACAGCAAAAT 1693
617 SGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACATTACAGCAATAAGATATTACTACACCGGCAATTAAC 1743
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnHisSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747ACTTGGAT..... 1755
667 AlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756ACCAAAAAAGAAATGGCC 1773
684 eHisSerLysGlyLeuThrThrIleTyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTGGCGAGAAAGATGCAACCAACCAAGCGGCGCT 1823

701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATGTGATTTACCAACCGGAGACCGGATCGCATTTACTGCTTCCG 1873
708 LSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGACCAAAATTTAAAC.....GGC 1893
724 roAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATACCCCAACA...AACGGCAACTGTTTTCACGCGGAGACCGCA 1940
741 AsnIleThrAlaThrLysGlyLysSerValPhePheAsp..... 753
1941 ACCGACGCGCTAC...AATCATTTAGAGCGGCGGTGCTCAAAATGAAG 1987
754IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAACAGACAGAAATCGTGGGACACAGATTGATCGACCGCACA 2037
764 IuleuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGGAAACTTC.....CATATTCAGGCGGAGCAAGC 2075
775GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGTTTCCCGCAATGTGCCAAAGTGAAGCGCATTTGCTTTACACA 2125
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn 799
2126 ATCAGCGCCAGACGATTTCGTGTCGACCGCATCAAGCCACACAATC 2175
799 799
2176 TGATACGCTTGGAGCTGACGGGTCTGACAACTTGATCCGAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACATTAAGTATGTTTCATTTGACAGACCGCATCAGAGCA 2275
806SerLeuArgGlnThrLysAspAspPheTr 815
2276 AATGACGCTTGCCGATGACGCTCATTTAAATCTCAGACGACTGCCACA 2325
815 TyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCACGCGCAATCTTAGTGCAGGCGGAGACACGCACTATACGTTACCG 2375
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAAGCGCACCAAAACGCAACCTCAGCCTCGTGGGCAATGCCAAGCA 2425
841GlnAsnSerSerSerIleThrLysAsn.....IleT 852
2426 CATTATATCAAGC.....ACATTAAAGCGACACATCGGCTCG 2466
852 hIleGluLysAlaIleAsnValThrLeuGluAlaIleAsnAlaAspIleLysG 868
2467 GACATGCTTCATTATCTTAAGCAACAAGCGGTCACAAACGCAAGTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTCCGAAACGCTAAGCAACGTAAGCAATCCGACGCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrLysLysAsnAlaAspIleLysG 899
2567 GCAATGTCCTCCAGCGCATAGCGATTTCCAT.....TTTAA 2607
899 LysAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCGCTTTACCGAAAAATCAGCGGCGGAGATGACGCAATTACA 2657

[illegible]

```

3453 AACCGGCGGGCTACACCGCTTTCCCGGGGCCCGCGCGCGG 3502
1131 ..... 1131
3503 ATTTGGCGCAACGACGACCCCAACCGCAACCGACGCGACGTG 3552
1132 .....AsnLysAspLe 1135
3553 ATACGC..CGTTATGCCAATACGGTTTAGTGAATTTCCGCCACGT 3599
1136 ThrSerLeuLysThrValAsnLierThrAlaSerGluLysValThrThr.T 1152
3600 CAACACCGTTTCCCGCTACGACGCAATTTGGACCCGGTGTTCGCAAG 3649
1152 hValaGlySerThrIleAsnAlaThrAsnLysAlaSerLierThrThr 1168
3650 ACCGCGCAACGCGGTTTGGACAAGCGGACATCCGGGACACCAACACTAC 3699
1169 LysThrGlyAspLieserGlyThrLieserGlyAsnThrValSerValse 1185
3700 CGTTGCC..AGATTTCCGCGCTACCGCCACCAACAACGACCTGGCCCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202
3747 AATCGGTATGCAAAAAACCTCGGACGCGGCGGTGCGCATCCTGTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
3797 CGCAACAACGCGGACCGGAACACCTTCGACGACGACATCGGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839 .....ACTCGGACAGCGTTGCCACCG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249
3861 TGCGGTTTTCGGGCAAT.....ACGACGATCGGACGAGTTGCATCGCA 3904
1249 euThrAlaThrGlyAsnThrLierThrThrGluAlaGlySerSerLierThr 1265
3905 TCAGCGCGGCGCGGCTTTAGTAGCGGACGCTTTCAGACGGCATCAGA 3954
1266 SerThrLysGlyGluValAspLeuAla.....GlnAsnGlySerI 1280
3955 GGCAAAATCCGGCGCGCGCTGCATTCAGCATTCAGGACAGATACCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CGCAGGTTTCGGCGGATTCGACATCCACGACATCGGCGCAACGCGCT 4054
1297 euThrThrValAlaGlySerAspLielLysAlaThrSerGlyThrLeuVal 1313
4055 ATTTCGTCACAA 4066
1314 IleAsnAlaLys 1317
seq_name: /cgn2_6/ptodate/1/1aa/5B_COMB.pep:US-08-617-697-4
seq_documentation_block:
? Sequence 4, Application US/08617697
? Patent No. 5977336
? GENERAL INFORMATION:
? APPLICANT: Barenkamp, Stephen J
? TITLE OF INVENTION: High Molecular Weight Surface Proteins
? TITLE OF INVENTION: Of No. 5977336-typeable Haemophilus
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Shoemaker and Maltare, Ltd.
? STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
? CITY: Arlington
? STATE: Virginia

```

```

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-4

```

```

alignment_scores:
  Quality: 274.50      Length: 1488
  Ratio: 0.419        Gaps: 73
  Percent Similarity: 44.019  Percent Identity: 19.220

```

```

alignment_block:
US-09-303-518D-653 x US-08-617-697-4

```

```

Align seq 1/1 to: US-08-617-697-4 from: 1 to: 1477

```

```

178 AATAAGGCAAGTTTGCAGTCGGGCGAAGATATTGAGTTTACACAA 227
    ||| ||| :::::||||| ||||| ::|||
199 AsnHisIleuIlethrvAlGly..LysAspIlySerVal..... 211
228 AAAAGGAGTGGTGGCGCAATGATGACGAAAGCCCGATGATTGATT 277
    ::|||::||| ::::: :::::
212 .....AsnIleuIleGlyIlyValIlyAsnGlu..... 221
278 TTTCGTGATGCGCTAACGGC.....GTGGGCGCATTTGGGCGCAT 321
    ::|||::||| ||||| ::::: |||||::|||
222 ..GlyValIleSerValAsnIlyGlySerIleSerIleuAlaGlyGln 237
322 CAA.....TATATTGTGAG 335
    ::|
238 LysIlethrvIleSerAspIleIleAsnProthrvIlethrvIleSerIleAl 254
336 CGTGCGCATRACGCGGCTATTAACATGTTGATTTGGTGGGAGGAA 385
    ::::: ::||| ::| ||||| ::|||::|||
254 AlaIleProGluAsnGluAlaValAsnIleuGlyAspIleheAlaIlySgl 271
386 GCAT.....CCGATCAGCACCGCTTTCT 411
    ::|||
271 LysAsnIleAsnValArgAlaIleAlaIlethrvIleArgAsnGlnGlyLysIleuSer 287
412 TACCAAAATGTGAAGAAATATATATTAAGACGAGACTAAGCGCATCC 461
    ::| ::|||::||| ::|||::|||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGGCGATTATCATATGCCGCTTTGACAAATTTGTACAGATG 511
    ::|||::||| ::|||

```

```

298 .....AsnIleValIleuSerA 303
512 CAGAACCTGTTGAG.....ATGACCACTTATATGAT 543
    ||::: |||
303 IAlYSGluGIyGluAlaIleGlyIyValIleSerAlaGlnAsnGln 319
544 GCGTGAAATACGCTGATTTAAATAATACCGTATGCTGTT.....CG 587
    ||| ::| ||||| ::|||
320 GlnAlaIySGlyIyLysIleuMetIlethrvGlyAspIlyValIleuLys 336
588 AATCGAGCA.....GCGACACAAATTT 610
    : |||||
336 SthrvGlyAlaValIleAspIleuSerGlyIySGluGlyIyValIleuLys 353
611 GCGGCTGATGAA...GACGAAACCAATTAAC..... 639
    ::|||::|||
353 euGIyGlyAspIleuArgGlyIyLysAsnGlyIleGlnIleuAlaIyS 369
640 .....CGCGAAAGTTCAATATGATTCAGAC..... 666
370 LysThrSerIleuGluIySGlySerThrIleAsnValSerGlyIySGluIy 386
667 .....GCATATCTTGG.....CTGCTGCTGACATA 694
386 sGIyGlyPheAlaIleValIlePrgIyAspIleAlaIleuIleAspIlyAsnI 403
695 CCTTGACAAATATGATCAGTGGTGGCAGATCACTTAGTAGAGCAA 744
    ||||| |||||
403 IeAsnAlaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGCTTTTACCACAGAGGCTCATTTGG 794
    ::|
410 .....AspIleAlaIySthrvGlyIyPheValGI 419
795 CGACAGTGGCTCACCATGTTATCTATGATGCCCAAGCAAAAGTGT 844
    ||||| ::|||::|||
419 uThrSerGlyHisAspIleuPheIleLysAsp..... 429
845 TAATTAATGGGTATTCGCAACAGCAACCCCTATATAGAAAAAGCAAT 894
    |||::|||::|||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCACAGTATGCTGAAGATTGTTCTATGATGAAATCTTTGCTGG 944
    |||::|||::|||
436 .....LysGluIleuLysAsp.....Ph 442
945 AGATACCCATTCAGTATTCCTAGAA...CCACATCAAAATGGGAAATACT 991
    ||||| ::|||
442 eaSpAsnValSerIleAsnAlaGluAspIleuPheAsnAsnIleGlyI 459
992 TTTTAAACGACATATATATGCGCGAGGAAAAATGATGCCAAACATAAA 1041
    ::::: ::|||::|||
459 IeAsnAspIleuPheProthrvGlyIleGlyIleAlaSerAspIleuLys 475
1042 CACTATTCCTACCTTATGATTTAAACACAGAACCGTTCATTTT... 1089
    ::|||
476 AsnSerGluIleuLysThrIleuLysAsnIleSerAsnIle 492
1090 .....AATGTTCTTATTCGAGACAGCAAGCAAGAAC 1120
    |||::|||
492 uIySAsnAlaIlePheIleAsnIleThrAlaSerGlyLysIleuThr.... 507
511 CTGTTATCATGCTGACGCTGCGTCAACAGTATTCAGCCAGACTGAT 1170
    |||||
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATTTCTTATTCAGCAAGGAAAGGATTTATACT 1220
    |||||
514 IleGlySerAsnSerHisIleuIleuHisSerIySgl..... 526
1221 TACACACACATCAACCAAGGCGGCGTGTGATTTGAGGAGTAAAT 1270
    ||| ::|||::|||
527 .....GlnArgIyGlyIyValGlnIleAspIyAspI 538

```


FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-625
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-719-641-4

alignment_scores:
Quality: 274.50 Length: 1488
Ratio: 0.419 Gaps: 73
Percent Similarity: 44.019 Percent Identity: 19.220

alignment block:

US-09-303-518d-653 x US-08-719-641-4 ..

Align seg 1/1 to: US-08-719-641-4 from: 1 to: 1477

```
178 AATAAGGAGAGTTGTCAGTGGGCGAAGAAATATTGAGTTTACACAA 227
    ||| ||| :||||| ||| ||| ||| ||| ||| ||| ||| |||
199 AsnHisGlyLeuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAGGAGAGTTGTCAGTGGGCGAAGAAATCGATGCAAGCCCGATGATG 277
    :||||| ||| :||| :||| :||| :||| :||| :||| :|||
212 .....AsnLeuIleGlyGlyValValLysAsnGlu..... 221
278 TTTCTGTGATCGCTTACGCGC.....GTGCGGCGATGCGCGCGAT 321
    :||||| ||| :||||| ||| ||| ||| ||| ||| ||| |||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyIn 237
322 CAA.....TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGCACATAACGCGCGCTATACATGTTGATTTTGGTGGCGAGGAA 385
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
254 AlaIleProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAT.....CCGATCAGCAGCGCTTTCT 411
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
271 LysAsnIleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATGTGAAAGAAATATATATAAGACGAGCTAACGCGCATCC 461
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
288 AlaAspSerValSerLysAsp....LysSerGly..... 297
462 TTATGCGCGCATTCATATCCGCGTTTGCACAAATTTGTCACAGATG 511
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
298 .....AsnIleValIleuSerA 303
512 CAGAACCTGTGAG.....ATGACCAATTATATGAT 543
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
303 LAlaSerGluGlyLysIleGluIleGlyValIleSerAlaGlnAsnGln 319
544 GGGTGAATACGCTGATTTAATAATACCTGATCGTGT.....CG 587
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
320 GlnAlaLysGlyLysLeuMetIleThrGlyAspLysValIleThrLeuLys 336
588 AATCGAGACA.....GGCAGACATATAT 610
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
336 SThrGlyAlaValIleAspLeuSerGlyLysGluGlyGluIleThrTyrL 353
611 GCGGCTGATGAA...GACGAACCCCAATAC..... 639
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
```

```
353 euGIyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGGAAAGTTCATATCATTTGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlyLys 386
667 .....GCATATTCCTTG.....CTGCGTGGTGGCAATA 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 sGIyGlyPheAlaIleValIleTrpGlyAspIleAlaLeuIleAspLysAsnI 403
695 CCTTGGCAAAATGATCAGTGGTGGCACAGTCAACTTAGTAGCGAA 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 LeasnAlaGln...GlySerGly..... 409
745 AAATTAACATAGCCCATATGTTTTCACACAGAGGCGTATTTGG 794
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTACCCAAATGTTATCTATGATGCCAAAGCAAAATGTGT 844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGCGTATTCGAAACAGCAACCCCTATATAGCAAAAGCAAT 894
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCCAGCTAGTTGTAAGATTGGTCTATGATGAAATCTTGTCTGG 944
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTTCTACGAA...CCACATCAAAATGGCAAAATACT 991
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459
992 TTTTAAACGACAAATATATATGCGCGAGAAATCGATCCCAACATATA 1041
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
459 LeasnAspGluPheProThrIleThrGlyAlaIleAsnAspProLysLys 475
1042 CACTATCTCTACCTTATAGATTAAACAGCAACCGCTCAATGTGT.. 1089
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnTyrLe 492
1090 .....AATGTTCTTTATCCGACAGCAAGCAAGAAC 1120
492 uLysAsnAlaIleThrPheMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGTCGAGGTGGGTCAACGTTATGCAACCCAGCATGAAT 1170
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATTTCTTATTGACAAAGAAAGGTAATGATTAAT 1220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACACAGCAACATCAACCAAGCGCGCGGCTTTGATTTGAGGTAAT 1270
    ||| :||| :||| :||| :||| :||| :||| :||| :|||
527 .....GlnArgGlyGlyValGlnIleAspLysAspI 538
1271 TTACGGTCTCGCTTAAAAACACAGAAACGTGGCAAGCGCGCGTTCAT 1320
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 IeThrSerLysGlyGlyAsnLeuThrIleThrSerGlyGlyIleTrpValAsp 554
1321 ATCAGTGTGCGCATACCGTTACTTGAAAGTAAACGCGGTGCAACGA 1370
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCACAAATCGGCAAGGACAGCTGCTGTTCAGACC..... 1413
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnSer 573
1414 .....AAAGGGAAACCAAGCGCTGCGTCAACGCTGGCGACGCTAA 1455
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
573 AlaIlePheGluGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
```

```
1456 GTCATCTTAGATCAGACGGCGAGATCAAGCAAAAAACAAGCTTAG 1505
      :::::
590 IleVal..... 591
1506 TGAATCGCTGGTGTACGGCGAGGGGACGTGCAACTGATCCGATA 1555
      :::::
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTCAACCCGCAAACTCATTTGGCTTGGCGGGGAGCGCTTG 1605
      :::::
602 Lys.....AspPheArgAlaAsnVal 610
1606 GATTGACGGGCAATTCGCTTGTCCACCGCATTCAAATACCGGTA 1655
      :::::
611 SerLeuasnGly.....ThrGly 617
1656 AGGGCGCATGATGTC.....AACCAATCAAGACAAGAAAT 1693
      :::::
617 SGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTCAGCAATAGATATTTACTACACCGGCAATTAAC 1743
      :::::
634 LyrThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerLyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 ....ACTTGGAT..... 1755
667 AlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPheIleLysTrpIleSerS 684
1756 .....AGCAAAAAAGAAATGGC 1773
684 eIAsnSerLysGlyLeuThrThrGlnLyrArgSerSerAlaGlyValAsn 700
1774 TACAACGGTTGGTTGGCGAAGAAATGCAACCAAGCAAGCGGGCGCT 1823
      :::::
701 PheasnGly.....ValasnGlyAsnMe 708
1824 CAATCTGAATTTACCAACCGGNAAGGAGTGCACATTACTGCTTCCG 1873
      :::::
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAATTTAAAC.....GGC 1893
724 roAsnGlnAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCAGCAACA...AACGGCAACTGTTTTCAGCGGAGACCGGAC 1940
      :::::
741 AsnIleThrAlaThrGlyGlySerValPhePheasp..... 753
1941 ACCGACGGCTAC...AATCATTTAGAACGGGGTGTCAAAAATGGAAG 1987
      :::::
754 .....IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCACAGAGGAATCGTGTGGGACACGATTGGATGACCGGACA 2037
      :::::
764 IuLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAACGGGAAACTTC.....CATATTCAGGCGGAGACAAGC 2075
      :::::
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCCGCAATGTGCCAAAGTGAAGCGGATGGCATTTAAGCA 2125
      :::::
789 apHe.....LysIleAsnLysAspLeuThrIleAsn 799
2126 ATCAGCGCCAAGAGTTTGGGTGTCCACCGGATCAAAACCAACAAATC 2175
799 ..... 799
```

```
2176 TGATACAGTTGCGAGCTGAGCGGGTCTGACAAGTTGTACGAAAAAACCAT 2225
      :::::
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACGATTAAGTATTGCTTCAATTGAGCAAGACGACATCAGAGCA 2275
      :::::
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCTTCGCGATCAGCTCATTTAAATCTCACAGACTTGCCACA 2325
      :::::
815 YrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCACGGCAATCTTAGTCAGGCGGAGACACGCACTATACGGTTACGGC 2375
      :::::
832 LeuGlyGlyAsnValThrLeuGlyLys..... 840
2376 CAACGCCACCAAAACGGCAACTCAGCTGTGGGCAATGCCAAGCA 2425
      :::::
841 .....GlnAsnSerSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAACGGCAACACATCGGCTTCG 2466
      :::::
852 hTrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAlaProAsn 868
2467 GACAAATGCTTCAATTATCTAAGCAACACCGCGTACAAAACGGCAGTCT 2516
      :::::
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCGCAACAGCTAAGGCAACGTAAAGCATTCGCCACATCAACG 2566
      :::::
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCATGTCTCCCTACCGCATAGCGATATCCAT.....TTTGA 2607
      :::::
899 LysAsnLeuThrIleSerGlnSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTACCGGAAAAATACGCGCGGCAAGATACGGCATTTCA 2657
      :::::
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrIleAlaGluIleAs 932
2658 CTTAAAGACACGAATGAGCGCTGCGGCGGCGACAGATTAAGCAATTT 2707
      :::::
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspLysAspL 949
2708 TAACTTGTACAACGCCACCATTTACATCAATTCGCCCTATGCACACGAT 2757
      :::::
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGGACGGCGCAACCGGACGTGCGGACATGCGCGCCGCCCGCTTC 2807
      :::::
956 .....LysArgAs 958
2808 GCGCGTTCCTCATTTATTCGTTACCGCGCAACTTCGCGACAGATCCGTT 2857
      :::::
958 nGlnArgSerIle..... 962
2858 TCAACACGCTGACGTTAAACGGCAATTGAACGCTCAGAGGAACATTCGCG 2907
      :::::
963 .....IleGlyGlyAspIleIleAsn..... 969
2908 TTTATGTGGAACTTTCGCTACCGGACGGCAAAATTTGAAGCTGGCGGA 2957
      :::::
970 .....LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTCC.....G 2965
      :::::
978 pSerAsnAsnAspAlaGluIleGlnIleGlyLysAsnIleSerGlnLysG 995
2966 AAGCACTTACACTTGGCT.....GTCAACAATACCGGCAACGAA 3006
      :::::
995 LylGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCCGTAACTCTGAGCAATTGACGGTA.....GTGAAGAAAAAGA 3047
```

```

1009 .....LysGlnIleThrIleLysLysGlyIleAspGlyIleAs 1021
1010 .....
1011 .....
1012 .....
1013 .....
1014 .....
1015 .....
1016 .....
1017 .....
1018 .....
1019 .....
1020 .....
1021 .....
1022 .....
1023 .....
1024 .....
1025 .....
1026 .....
1027 .....
1028 .....
1029 .....
1030 .....
1031 .....
1032 .....
1033 .....
1034 .....
1035 .....
1036 .....
1037 .....
1038 .....
1039 .....
1040 .....
1041 .....
1042 .....
1043 .....
1044 .....
1045 .....
1046 .....
1047 .....
1048 .....
1049 .....
1050 .....
1051 .....
1052 .....
1053 .....
1054 .....
1055 .....
1056 .....
1057 .....
1058 .....
1059 .....
1060 .....
1061 .....
1062 .....
1063 .....
1064 .....
1065 .....
1066 .....
1067 .....
1068 .....
1069 .....
1070 .....
1071 .....
1072 .....
1073 .....
1074 .....
1075 .....
1076 .....
1077 .....
1078 .....
1079 .....
1080 .....
1081 .....
1082 .....
1083 .....
1084 .....
1085 .....
1086 .....
1087 .....
1088 .....
1089 .....
1090 .....
1091 .....
1092 .....
1093 .....
1094 .....
1095 .....
1096 .....
1097 .....
1098 .....
1099 .....
1100 .....
1101 .....
1102 .....
1103 .....
1104 .....
1105 .....
1106 .....
1107 .....
1108 .....
1109 .....
1110 .....
1111 .....
1112 .....
1113 .....
1114 .....
1115 .....
1116 .....
1117 .....
1118 .....
1119 .....
1120 .....
1121 .....
1122 .....
1123 .....
1124 .....
1125 .....
1126 .....
1127 .....
1128 .....
1129 .....
1130 .....
1131 .....
1132 .....
1133 .....
1134 .....
1135 .....
1136 .....
1137 .....
1138 .....
1139 .....
1140 .....
1141 .....
1142 .....
1143 .....
1144 .....
1145 .....
1146 .....
1147 .....
1148 .....
1149 .....
1150 .....
1151 .....
1152 .....
1153 .....
1154 .....
1155 .....
1156 .....
1157 .....
1158 .....
1159 .....
1160 .....
1161 .....
1162 .....
1163 .....
1164 .....
1165 .....
1166 .....
1167 .....
1168 .....
1169 .....
1170 .....
1171 .....
1172 .....
1173 .....
1174 .....
1175 .....
1176 .....
1177 .....
1178 .....
1179 .....
1180 .....
1181 .....
1182 .....
1183 .....
1184 .....
1185 .....
1186 .....
1187 .....
1188 .....
1189 .....
1190 .....
1191 .....
1192 .....
1193 .....
1194 .....
1195 .....
1196 .....
1197 .....
1198 .....
1199 .....
1200 .....
1201 .....
1202 .....
1203 .....
1204 .....
1205 .....
1206 .....
1207 .....
1208 .....
1209 .....
1210 .....
1211 .....
1212 .....
1213 .....
1214 .....
1215 .....
1216 .....
1217 .....
1218 .....
1219 .....
1220 .....
1221 .....
1222 .....
1223 .....
1224 .....
1225 .....
1226 .....
1227 .....
1228 .....
1229 .....
1230 .....
1231 .....
1232 .....
1233 .....
1234 .....
1235 .....
1236 .....
1237 .....
1238 .....
1239 .....
1240 .....
1241 .....
1242 .....
1243 .....
1244 .....
1245 .....
1246 .....
1247 .....
1248 .....
1249 .....
1250 .....
1251 .....
1252 .....
1253 .....
1254 .....
1255 .....
1256 .....
1257 .....
1258 .....
1259 .....
1260 .....
1261 .....
1262 .....
1263 .....
1264 .....
1265 .....
1266 .....
1267 .....
1268 .....
1269 .....
1270 .....
1271 .....
1272 .....
1273 .....
1274 .....
1275 .....
1276 .....
1277 .....
1278 .....
1279 .....
1280 .....
1281 .....
1282 .....
1283 .....
1284 .....
1285 .....
1286 .....
1287 .....
1288 .....
1289 .....
1290 .....
1291 .....
1292 .....
1293 .....
1294 .....
1295 .....
1296 .....
1297 .....
1298 .....
1299 .....
1300 .....
1301 .....
1302 .....
1303 .....
1304 .....
1305 .....
1306 .....
1307 .....
1308 .....
1309 .....
1310 .....
1311 .....
1312 .....
1313 .....
1314 .....
1315 .....
1316 .....
1317 .....
1318 .....
1319 .....
1320 .....
1321 .....
1322 .....
1323 .....
1324 .....
1325 .....
1326 .....
1327 .....
1328 .....
1329 .....
1330 .....
1331 .....
1332 .....
1333 .....
1334 .....
1335 .....
1336 .....
1337 .....
1338 .....
1339 .....
1340 .....
1341 .....
1342 .....
1343 .....
1344 .....
1345 .....
1346 .....
1347 .....
1348 .....
1349 .....
1350 .....
1351 .....
1352 .....
1353 .....
1354 .....
1355 .....
1356 .....
1357 .....
1358 .....
1359 .....
1360 .....
1361 .....
1362 .....
1363 .....
1364 .....
1365 .....
1366 .....
1367 .....
1368 .....
1369 .....
1370 .....
1371 .....
1372 .....
1373 .....
1374 .....
1375 .....
1376 .....
1377 .....
1378 .....
1379 .....
1380 .....
1381 .....
1382 .....
1383 .....
1384 .....
1385 .....
1386 .....
1387 .....
1388 .....
1389 .....
1390 .....
1391 .....
1392 .....
1393 .....
1394 .....
1395 .....
1396 .....
1397 .....
1398 .....
1399 .....
1400 .....
1401 .....
1402 .....
1403 .....
1404 .....
1405 .....
1406 .....
1407 .....
1408 .....
1409 .....
1410 .....
1411 .....
1412 .....
1413 .....
1414 .....
1415 .....
1416 .....
1417 .....
1418 .....
1419 .....
1420 .....
1421 .....
1422 .....
1423 .....
1424 .....
1425 .....
1426 .....
1427 .....
1428 .....
1429 .....
1430 .....
1431 .....
1432 .....
1433 .....
1434 .....
1435 .....
1436 .....
1437 .....
1438 .....
1439 .....
1440 .....
1441 .....
1442 .....
1443 .....
1444 .....
1445 .....
1446 .....
1447 .....
1448 .....
1449 .....
1450 .....
1451 .....
1452 .....
1453 .....
1454 .....
1455 .....
1456 .....
1457 .....
1458 .....
1459 .....
1460 .....
1461 .....
1462 .....
1463 .....
1464 .....
1465 .....
1466 .....
1467 .....
1468 .....
1469 .....
1470 .....
1471 .....
1472 .....
1473 .....
1474 .....
1475 .....
1476 .....
1477 .....
1478 .....
1479 .....
1480 .....
1481 .....
1482 .....
1483 .....
1484 .....
1485 .....
1486 .....
1487 .....
1488 .....
1489 .....
1490 .....
1491 .....
1492 .....
1493 .....
1494 .....
1495 .....
1496 .....
1497 .....
1498 .....
1499 .....
1500 .....

```

```

1232 uthrValIglYasnGlyAlaGluIleAsnAlaThrGluIglYAlaIleThrL 1249
1233 .....
1234 .....
1235 .....
1236 .....
1237 .....
1238 .....
1239 .....
1240 .....
1241 .....
1242 .....
1243 .....
1244 .....
1245 .....
1246 .....
1247 .....
1248 .....
1249 .....
1250 .....
1251 .....
1252 .....
1253 .....
1254 .....
1255 .....
1256 .....
1257 .....
1258 .....
1259 .....
1260 .....
1261 .....
1262 .....
1263 .....
1264 .....
1265 .....
1266 .....
1267 .....
1268 .....
1269 .....
1270 .....
1271 .....
1272 .....
1273 .....
1274 .....
1275 .....
1276 .....
1277 .....
1278 .....
1279 .....
1280 .....
1281 .....
1282 .....
1283 .....
1284 .....
1285 .....
1286 .....
1287 .....
1288 .....
1289 .....
1290 .....
1291 .....
1292 .....
1293 .....
1294 .....
1295 .....
1296 .....
1297 .....
1298 .....
1299 .....
1300 .....
1301 .....
1302 .....
1303 .....
1304 .....
1305 .....
1306 .....
1307 .....
1308 .....
1309 .....
1310 .....
1311 .....
1312 .....
1313 .....
1314 .....
1315 .....
1316 .....
1317 .....
1318 .....
1319 .....
1320 .....
1321 .....
1322 .....
1323 .....
1324 .....
1325 .....
1326 .....
1327 .....
1328 .....
1329 .....
1330 .....
1331 .....
1332 .....
1333 .....
1334 .....
1335 .....
1336 .....
1337 .....
1338 .....
1339 .....
1340 .....
1341 .....
1342 .....
1343 .....
1344 .....
1345 .....
1346 .....
1347 .....
1348 .....
1349 .....
1350 .....
1351 .....
1352 .....
1353 .....
1354 .....
1355 .....
1356 .....
1357 .....
1358 .....
1359 .....
1360 .....
1361 .....
1362 .....
1363 .....
1364 .....
1365 .....
1366 .....
1367 .....
1368 .....
1369 .....
1370 .....
1371 .....
1372 .....
1373 .....
1374 .....
1375 .....
1376 .....
1377 .....
1378 .....
1379 .....
1380 .....
1381 .....
1382 .....
1383 .....
1384 .....
1385 .....
1386 .....
1387 .....
1388 .....
1389 .....
1390 .....
1391 .....
1392 .....
1393 .....
1394 .....
1395 .....
1396 .....
1397 .....
1398 .....
1399 .....
1400 .....
1401 .....
1402 .....
1403 .....
1404 .....
1405 .....
1406 .....
1407 .....
1408 .....
1409 .....
1410 .....
1411 .....
1412 .....
1413 .....
1414 .....
1415 .....
1416 .....
1417 .....
1418 .....
1419 .....
1420 .....
1421 .....
1422 .....
1423 .....
1424 .....
1425 .....
1426 .....
1427 .....
1428 .....
1429 .....
1430 .....
1431 .....
1432 .....
1433 .....
1434 .....
1435 .....
1436 .....
1437 .....
1438 .....
1439 .....
1440 .....
1441 .....
1442 .....
1443 .....
1444 .....
1445 .....
1446 .....
1447 .....
1448 .....
1449 .....
1450 .....
1451 .....
1452 .....
1453 .....
1454 .....
1455 .....
1456 .....
1457 .....
1458 .....
1459 .....
1460 .....
1461 .....
1462 .....
1463 .....
1464 .....
1465 .....
1466 .....
1467 .....
1468 .....
1469 .....
1470 .....
1471 .....
1472 .....
1473 .....
1474 .....
1475 .....
1476 .....
1477 .....
1478 .....
1479 .....
1480 .....
1481 .....
1482 .....
1483 .....
1484 .....
1485 .....
1486 .....
1487 .....
1488 .....
1489 .....
1490 .....
1491 .....
1492 .....
1493 .....
1494 .....
1495 .....
1496 .....
1497 .....
1498 .....
1499 .....
1500 .....

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-460-269C-4

seq_documentation_block:

Sequence 4, Application US/08460269C

Patent No. 6197548

GENERAL INFORMATION:

APPLICANT: CLARE, JEFFREY J.

ROMANOS, MICHAEL A.

TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Millen, White, Zelmano & Branigan, P.C.

STREET: 2200 Clarendon Blvd., Suite 1400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,269C

FILING DATE: 02-Jun-1995

ATTORNEY/AGENT INFORMATION:

NAME: Lebovitz, Richard M.

REGISTRATION NUMBER: 37,067

REFERENCE/DOCKET NUMBER: Popov-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 243-6333

TELEFAX: (703) 243-6410

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 911 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-08-460-269C-4

alignment_scores:

Quality: 274.00

Ratio: 0.584

Percent Similarity: 40.818

Percent Identity: 20.540

Length: 1149

Gaps: 50

alignment_block:

US-09-303-518D-653 x US-08-460-269C-4 ..

Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911

```

1306 GCGCGCGCGCTCATATCAGTGTGCGCAGTACCGCTTACTTGCAAAATATA 1355
      |||||
      58 GlyAlaGlyValArgThrAlaThrGlyThrIle.....LysValSer 72
1356 CGCGCGGCGAAACGACCGCGCTGCAAAATCGGCAAGGACGACGCTCTGG 1405
      |||||
      72 rGly.....ArgGlnAlaGlnGlyValLeuLeuG 82
1406 TTCAAGCCAAAGGGGAAACCAAGGCTCGGTACGCTGGCGGCGAGGTAA 1455
      :|||
      82 IuAsnProAlaAlaGlu.....LeuArgPheGlnAsnGlySer 94
1456 GTATCTTGAATCAGCAGCGGCGGCGATCAAGCAAAACCAAGCCCTTAG 1505
      |||
      95 ValThrSerSerGlyLeuPheAspGlyGlyValArgArgPheLeu... 110
1506 TGAATCGCTGTGTGACGCGGCGGACGCGTCAACTGTAATGCGGATA 1555
      |||||
      111 .....GlyThrValThrValLysAla.... 117
1556 ATCAGTTCAACCCCGACAACTATTTGGCTTTCGCGGCGAGCTTTG 1605
      |||
      118 .....GlyLysLeu 120
1606 GATTTGACGGGCGATTCGCTTGTTCACCGCATTCMAAATACGATGA 1655
      :|||
      121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspSer 137
1656 AGGGCGCATATTTGTCAACCAATCAAGCAAGAAATCCACCGTT... 1701
      |||||
      137 pGlyIleAlaLeuThrValAlaGlyGlnAlaGlnAlaSerIleAla 154
1702 ....ACCATTACAGCAATAAAGATATTACTACACCGGCATTAACAC 1746
      |||||
      154 spSerThrLeuGlnGlyAlaGlyValArgValArgGlyAlaAsn 170
1747 AACTGTGATGCAAAAGAAATTCCTACACGCTTGTGGCGGAGAA 1796
      :|||
      171 ValThrValGlnArg..... 175
1797 AGATGCAACCAAAACGAGCGGCGCTCAATCTGAAT...T 1834
      :|||
      176 ....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnPro 191
1835 ACCACCGGGAAGAGCGGATCGCACTTACTGCTTCCGCGGACAAAT 1884
      |||||
      191 euGlnProGlyAspLeuProProSerArgValIleLeuGlyAspThrSer 207
1885 TTAAAGGCAATATCAGCAAAACGCAAACTGTTTTCAGCGGCGAG 1934
      :|||
      208 ValThrIle.....ValProAlaSerGly..... 215
1935 ACCGACACCGCAGCGCTTACATATTAGGAAGCGGTGCTCAAAATGG 1984
      :|||
      216 ....AlaProAlaAlaValSerValPheGlyAlaAsnGluLeuThrVal 231
1985 AAGGTATCCCAAGAGAAATCGTGGGACAAACGATTGGATGACCGC 2034
      :|||
      231 spGly..... 232
2035 ACATTTAAGCGGAAACTTCATATTACGCGGACAGCGGCTTTC 2084
      |||||
      233 .....GlyHisIleThrGlyValArgAlaAla.... 241
2085 CCGCAATGTTCGCAAGAGCGGATGGCATTTAAGCAATCAGCGCC 2134
      :|||
      242 ....GlyValAlaAlaIleMetAspGly..... 248
2135 AAGCATTTTCGTGTGCGACCGCATCAAAACGACACATCTGTACAGT 2184
      :|||
      248 ..... 248

```

```

2185 TCGACCTGAGCGGCTCTGACAAAGTTGTACGCAAAACCAATACCGAGA 2234
      :|||
      248 ..... 248
2235 TAAAGTATGCTTCATTAGCAAGACCGCATCAGAGCGCAATGTACGC 2284
      :|||
      249 ....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259
2285 TTGCGGATCAGGCTCATTTAATCTCACAGGACTTGGCACATCAGCGC 2334
      :|||
      260 .....ArgGly 261
2335 AATCTTAGCGGCGGAGACCGCAGCATATACGCTTACGCGCAAGCGCAC 2384
      :|||
      262 AspAlaProAlaGlyGly.....AlaValProGlyGlyAlaVal 274
2385 CCAAAACGCG.....A 2395
      :|||
      274 LProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspVal 291
2396 ACCCTGAGCTGTGGGCAATGCGGCAACATTTAATCAAGCCACATTA 2445
      :|||
      291 erAspSerThrValAspLeuAlaGlnSerIleValGlnAlaProGlnLeu 307
308 GlyAlaAlaIleArgAlaGlyArgGlyValArgValThrValSerGly 324
2491 .....AACACGCGGTACAAACGCG..... 2511
      :|||
      324 YserLeuSerAlaProHisGlyAsnValIleGlnThrGlyGlyAla 341
2512 .....AGTGTACGCTTCCGACAC 2532
      :|||
      341 rArgPheProProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGGCAAC.....GT 2546
      :|||
      358 AlaArgAlaGlnGlyArgAlaLeuLeuThrArgValLeuProGluProVa 374
2547 AAGCCATTCGCGCATCAACGCGAATGCTCCCTAGCCGATTAAGCAGTAT 2596
      :|||
      374 LysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnLysP.....IleV 389
2597 TCCATTTGAAACAGCGCTTACCGGAAATCAAGCGGCGGAGGAT 2646
      :|||
      389 AlaIleThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGATTTACCTTAAAGACAGCGAATGACGCTGCGTGGCGACGGA 2696
      :|||
      406 ValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAl 420
2697 ATTAGCAATTTAAACCTTGACAAGCCACCATTTACATCAATTCGGCT 2746
      :|||
      420 aValAspSerLeuSerIleAspAsnAlaThrTrpAlaMetThrAsp... 435
2747 ATCGACACGATCGGCGAGCGCGCAACCGGACGAGCGCGAGATGCGCG 2796
      :|||
      436 .....AsnSerAsnValGlyAlaLeu 442
2797 CGCGCGCTTTCGCGCGCTTCCTATTATCGTTACGCGGCAACTTCGCG 2846
      :|||
      443 ArgLeuAlaSerAspGlySer...ValAspPheGlnGlnProAlaGlnAl 458
2847 AGAATCCGCTTTCACACGCTGACGCTAAACGCGCAATTTGAACGTCAGG 2896
      :|||
      458 a...GlyArgPheLysCysLeuMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCGCTTATGTGCGAATCTTCGCTACGCGGCGGCAATTTG 2946
      :|||
      473 LLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeu 489

```



```

2947 AAGCTGGCGAAAAGTTCGAGGACATTACACCTTGGCTGTCAACATAC 2996
      ::::::::::::::: ||| :::::::::::::::
490 ValValMetArgAspAlaSerGlnHisArgLeuLeuValArgAsnSe 506
      ::::::::::::::: ||| :::::::::::::::
2997 CGGCAACGAAACCGGTAACTCTGAGCAATTGACGATAGTGAAGAAAG 3046
      ::::::::::::::: ||| :::::::::::::::
506 rGlySerGlnProAlaSerGlyAsnThrMetLeuLeuValGln..... 520
3047 ACAACACACCGCTGTCCGAAATCTTAATTTCACCTCGCAAAACGAA... 3093
      ||||| ::::::::::::::: ||||| |||||
521 .....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnGlySAsp 535
3094 ...CACGTGATGCGCGCATGGCGTATATCAGCTTATCCGCAAA...GA 3137
      ||||| ||||| ::::::::::::::: ||||| :::::::::::::::
536 GilyValAlaSplGlyThrArgGlyArgGlyLeuAlaAlaAsnGlyAs 552
3138 CGGCGAGTTCGGCTGCAATACCGGTCAAAAGACAGAGCTTTCCGACA 3187
      ::::::::::::::: ||| :::::::::::::::
552 nGlyGlnTyrSerLeu.Val..... 558
3188 AACTCGGAGGGGGAGAAACAGAGCCGCTTGACGGCAAAACAGGCA 3237
558 ..... 558
3238 CAACCTGCCGCAACACAGCGGAGAAAGAACACGCGCAAAAGCTTGA 3287
558 ..... 558
3288 CGCGCTGATTCGGCGCGGCGCATGCCACGAAAGCGAAGATGTTG 3337
      ||||| ||||| |||||
559 .....GlyAlaIysAlaProAlaProAlaProAlaProAla 569
3338 CCGAACCGGCGCGGAGGAGGAGGAAATCGCGCATATGACGGGG 3387
      ||| ::::::::::::::: |||
570 Pro...GlnProGlyProGln..... 575
3388 GAGGAGAGAAAAAACGGGTGACGGGATTAAGACCGCCTTGCGGAA 3437
575 ..... 575
3438 ACAGCCGGAAGCGGAACCCGCGGCTACACCGCCTTCCCGCGGCC 3487
      ||||| ||||| |||||
576 .....ProGlyProGlnProProGln.ProProGlnPr 586
3488 GCCGCGCCCGCGGATTTGCCGAACCGCCCAACCGGACCGCA 3537
      ||||| ||||| |||||
586 oProGlnProProGlnArgGlnProGlnAla.ProAlaProGlnProPro 602
3538 CCGCAGCGGACCTGATCAACCGCTTATGCCAATAGCGGTTGAGTGAAT 3587
      ||||| ||||| |||||
603 AlAGlyArgGlyLeu...SerAlaAlaAlaAsnAlaValAlaThrG1 618
3588 TTCCGCCAGCGTCAACAGGCTT.....TTCCGCCGTACAGAGCAATTGG 3631
      ::::::::::::::: ||| ::::::::::::::: |||||
618 yGlyValGlyLeuAlaSerThrLeuTyrAlaGlnSerAsnAlaLeuS 635
3632 ACCGCGTGTTCGCAAGACCGCGCAAC.....GCCGTTGG 3669
      ::::::::::::::: ||| |||||
635 eRlySArgLeuGlyGlyLeuAlaArgLeuAsnProAspAlaGlyAlaTTP 651
3670 ACAAGCGGATCCGGGACACCAACACTACGTTCCGAGATTTCCGCGC 3719
      ||||| ::::::::::::::: |||
652 GlyArgGlyPhe.....Al 656
3720 CTACCGCCCAACACGACCTGGCCAAATCGGTATGCAAAAAACCTTG 3769
      ||||| ||||| |||||
656 aGlnaArgGlnGlnLeuAsp..... 662
3770 GCAGCGGGGGCGTGGCATCTGTGTTTCGACACACGCGAGCAACACC 3819
      ||||| ::::::::::::::: |||||
663 .....AsnArgAlaGlyArgGlyArg 668
3820 TTTCAGCAGAC.....GGCATCGGCACTCGGCACG 3848

```

```

||||| ::::::::::::::: ||| :::::::::::::::
669 PheAspGlnIysValAlaGlyPheGlnLeuGlyAlaAspHisAlaValAl 685
3849 GCTTCCCGCCACCGGTGCC.....GTTTCGGGCAATACGGCATCGGAGCT 3892
      ::::::::::::::: ||| ::::::::::::::: |||||
685 aValAlaGlyIysArgTyrPheHisLeuGlyGlyLeuAlaGlyThrArg 702
3893 TCGACATCGCATCGACCGCGCGGCTTTTATGATCGGACGCTT... 3939
      ||| ||||| ::::::::::::::: |||||
702 IysAspArgGlyPheThrGlnAspGlyGlyGlyGlnHisThrAspSerValHis 718
3939 ..... 3939
719 ValGlyGlyTyrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAl 735
3940 .....TCAG 3943
735 aThrLeuArgAlaSerArgLeuGlnAsnAspPheIysValAlaGlySerA 752
3944 ACGGC.....ATCAGAGGCAAAATCCGCCCGCGCTGCTCATTCAGCG 3987
      ||||| ::::::::::::::: ||| :::::::::::::::
752 SPGlyTyrAlaValIysGlyIysTyrArgThr.....HisGly 764
3988 ATTCAGGCAAGATACCGCGCAGT.....TTCCGGCGGATT 4022
      ::| ||| ::::::::::::::: |||||
765 ValGlyAlaSerLeuGlnAlaGlyArgArgPheAlaHisAlaAspGlyTr 781
4023 CGGCATCGAACCGCACATCGCGCGCATATTTGCTCCAAAACCGG 4072
      ::| ||||| ::::::::::::::: ||| :::::::::::::::
781 PheLeuGlnProGlnIleAlaGlyLeuAlaValPheArgValGlyIys 798
4073 ATTACCGATACGAAAC...GTCAATATCGGCACCGCGGCTTGCAATTC 4119
      ::| ||||| ||| ::::::::::::::: |||
798 eTyrArgAlaAlaAsnGlyLeuArgValArgAspGlyGlySer 814
4120 AACCGCTACCGCGCGGATTAGGACAGATTATTCATCAACCGGCGCA 4169
      ||| ||||| ::::::::::::::: |||
815 ValLeuGlyArgLeuGlyLeuGlyValGlyIysArgIleGlyLeuAlaG1 831
4170 ACACATTTCCATCAACGCTTATTTGAGCTGTCTATACCATCCGCTT 4219
      ::| ::::::::::::::: ||| |||||
831 yGlyArgGlnValGlnProTyrIleIysAlaSerValLeuGlnIlePheA 848
4220 CCGGCAAGATCCGAACCGCGCTCAATACCGCGCTTATGCGCAGATTC 4269
      ::| ||| ::::::::::::::: ||| :::::::::::::::
848 SPGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGlyLeu 864
4270 GGCAAAACCGCAGTGGCGAATGGGCGGTAAACGCCGAATCAAAAGTTT 4319
      ||||| ||||| |||
865 ArgGlyThrArg...AlaGlyLeuGly..... 872
4320 CACGCTGTCCCTCCAGCGTGGCGCGCCGCAAGGGCGCCCAATTGGAAGCG 4368
      ||||| ::::::::::::::: |||||
873 ...LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaS 888
4369 .....CACACACGCGGCGCATCAAAATTAAGCTCCGCTGG 4404
      ::::::::::::::: ||| |||||
888 eTyrGlnTyrSerIysGlyProIysLeuAlaMetProTTP 901

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:us-08-460-269C-2
seq_documentation block:
; Sequence 2, Application us/08460269C
; Patent No. 6197548
; GENERAL INFORMATION:
; APPLICANT: CLARE, JEFFREY J.
; ROMANOS, MICHAEL A.
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
; STREET: 2200 Clarendon Blvd., Suite 1400
;

```

```

CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6410
TELEFAX: (703) 243-6333
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 910 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2

```

```

alignment_scores:
Quality: 268.00      Length: 1156
Ratio: 0.551         Gaps: 53
Percent Similarity: 42.042   Percent Identity: 21.367

```

alignment_block:

US-09-303-518D-653 x US-08-460-269C-2 ..

Align seg 1/1 to: US-08-460-269C-2 from: 1 to: 910

```

1366 AAGGACCGCCTGTCCAAATCGGCAAGG.....ACGCT 1400
      ||| |||||:||||| |||||: |||
      2 AsnMetSerLeuSerArgIleValIysAlaIaProLeuArgArgThr 18
1401 GGTGTTCAAGCCCAAGG..... 1419
      |||:||||| |||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaIaProAlaAlaHisAla 35
1420 ....GAAACCAAGGCTCGTCAGCGTGCGAGCGTAAGTCATCTTA 1464
      :|||:|||||: |||:|||||: |||: |||: |||: |||: |||
      35 sPTrpAsnAsnGlnSerIleValIysThrGlyIuArgIuGlnHisGlyIle 51
1465 GATCAGCAGCGGAGCATCAAGCAAAAAACAAGCCTTAGTCAATCGG 1514
      |||:||||| |||
      52 HisIleGlnIleYserAspProGlyIyAlaArgThrAlaSerGlyThr 68
1515 CTGG...GTGAGCGGAGGAGGAGGTGCAACTGATGCCGATATACAGT 1561
      :|||:|||||: |||: |||: |||: |||: |||: |||: |||
      68 rIleIysValSerGlyArgGlnAlaGlnGlyIleLeuLeuGluAsn... 83
1562 TCACCCCGCAAACTGATTTTCGCTTCGCGGCGGACGCTTGATTTG 1611
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      84 .....ProAlaIaGluLeuGlnPheArgAsnGlySerValThrSer 97
1612 AAGGGGATTCGCTTTCGTCACCGCATTCGAAATACCGATGAGGAGG 1661
      ||| |||: |||: |||: |||: |||: |||: |||: |||
      98 SerGly.....GlnLeuSerAspSpGlyIle 106
1662 GATGATGTCACCAATCAAGACAAGAAATCCACGTTTACCATACAG 1711
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      106 e.....ArgArgPheLeuGlyThrValThrValIysAla 117
1712 GCATATAA.....GATATTACTACAAACGGCAATACAAACAACCTTG 1752
      :|||: |||: |||: |||: |||: |||: |||: |||: |||

```

```

117 laGlyLysLeuValAlaAspHisAlaThrLeuAlaAsnValGlyAspThr 133
1753 GATACCAAAAAAGAAATTGCCACACAGCGTGGTTGGCGAGAAAGATGC 1802
      ||| |||: |||: |||: |||: |||: |||: |||: |||
      134 .....TrpAspAspGlyIleAl 140
1803 AACCAAAACGAACGGCGGCTCATCTGAAATACCAACGGGAAGAGCG 1852
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      140 aLeuIyValAlaGlyGlu.....GlnAlaGlnIaSerIleAla 154
1853 ATCGCACTTTCCTGCTTCGCGGAGACAAATTTA.....AACGGCAAT 1896
      ||| ||||| |||: |||: |||: |||: |||: |||: |||: |||
      154 SerThrLeuGlnGlyAlaGlyValGlnIleGlyArgGlyAlaAsn 170
1897 ATCAGCAACAAACGGCAACTGTTTTCAGCGGACCGACACCGCA 1946
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      171 ValThrValGlnArgSerAlaIleValAspIyGlyLeu..... 183
1947 CGCCTACAAATCATTTAGAGACGGGTGCTCA...AAATGAGAGTATCC 1993
      |||: |||: |||: |||: |||: |||: |||: |||: |||
      184 .....HisIleGlyAlaLeuGlnSerLeuGlnProGluAspLeuP 197
1994 CACAAGAGAAATCGTGTGGACAAACGATGTGATCGACCGACA..... 2037
      |||: |||: |||: |||: |||: |||: |||: |||: |||
      197 rOProSerArgValValLeuArgAspThrAsnValThrAlaValProAla 213
2038 .....TTTAAAGC 2045
214 SerGlyAlaProAlaIaValSerValLeuGlyAlaSerGlyLeuThrLe 230
2046 GGAACACTTCATATTCAGGCGGACAGCGGTGTTCCCGCAATGTTG 2095
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      230 uAspIyGlyHisIleThrGlyGlyArgAlaIa.....GlyAla 244
2096 CCAAAAGTGAAGCGATTGCGATTACCAATCAAGCCCAACAGCTTTC 2145
      ||| |||: |||: |||: |||: |||: |||: |||: |||
      244 la.....AlaMetGln 247
2146 GGTGTGCGACCGCATCAAGCGCACACATCTGTACCTTGAGAC.... 2190
      |||: |||: |||: |||: |||: |||: |||: |||: |||
      248 GlyAlaValIaHisLeuGlnArgAlaThrIleArgArgGlyAspAlaLe 264
2190 ..... 2190
264 uAlaGlyIyAlaValaIProGlyIyAlaValaIProGlyIyAlaValaIProG 281
2191 .....TGACGGGT 2199
281 IyGlyPheGlyProGlyIyPheGlyProValLeuAspGlyTrpIyGly 297
2200 CTGACAAGTTGTACGGAAGAAACCATTTACGACGATAAGTATGCTTC 2249
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      298 Val..... 298
2250 ATTGACAAGACCGACATCAGAGGC...AATGTACGCTTCGCGATCAG 2296
      |||: |||: |||: |||: |||: |||: |||: |||: |||
      299 .....AspValSerGlySerSerValGluLeuAlaGln... 309
2297 CTCATTTAAATCTCACAGGACTTGCACACATCAACGGCAATCTTAGTCA 2346
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      310 .....SerIleValGlnAlaProGlyLeuGlyAlaIaIleArgVal 323
2347 GGCGGAGACAGCACTATAGGTTACGGGCAAC.....GCCACCCAAA 2390
      ||| |||: |||: |||: |||: |||: |||: |||: |||
      324 GlyArgGlyAlaArgValaThrValaProGlyIySerLeuSerIleuAlaProH 340
2391 CGGCAACCTCAGCCTCGTGGGCAATGCCCAACAATTTATATCAACGC. 2439
      :|||: |||: |||: |||: |||: |||: |||: |||: |||
      340 sGlyAsnValIleGlyThrGlyIyAlaArgArgPheAlaProGlnAla 357
2440 .....ACATTAACGGCAACATGAGCTTCGAGCAATGCT 2475
      |||: |||: |||: |||: |||: |||: |||: |||: |||
      357 laProLeuSerIleThrLeuGlnAlaGlyAlaHisAlaGlnGlyAla 373

```



```

      805 ArgValArgAspIuGIyGlySerSerValIleuGIyArgIleuGIyLeuGI 821
4143 GGCAGATTATTCATCAACCGCGCAACACATTCATCAGCCCTTATT 4192
      821 uValGIyArgIleGIleuGIleuGIyArgIyValGIleuGIyLeuGI 838
4193 TGAGCCTGTCTTATACCGATCCGCTTCGCGCAAAAGTCGCAAGCGCGGC 4242
      838 leuValSerValIleuGIleuGIleuPheAspIyAlaGIyThrValIleuGI 854
4243 AATACCGCCCTATTGGCGAGGATTTCGCGCAAAACCGCAGTGGCGAATG 4292
      855 AsnGIyIleAlaHisArgThrGIleuArgGIyThrArg...AlaGIleu 870
4293 GGGCGTAAACGCCGAATCAAGGTTTCACGCTGTCGCCACGCGTGGCG 4342
      870 uGIy...LeuGIyMetAlaAlaAla 878
4343 CCGCCAAAGGCGCGCAATTGGAACG...CAGCACAGCGCGGCGATC 4386
      878 euGIyArgGIyHisSerLeuTyAlaSerTyGIyTyrSerGIyLeuGIyPro 894
4387 AATATGCGTACCGCTGG 4404
      895 LysLeuAlaMetProTyr 900

seq_name: /cgn2_6/plodata/1/1aa/5A_COMB.pep:us-08-038-682-2
seq_documentation_block:
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-038-682-2
alignment_scores:

```

```

      Quality: 265.00      Length: 1297
      Ratio: 0.438      Gaps: 73
      Percent Similarity: 46.646      Percent Identity: 21.126
alignment_block:
Align seg 1/1 to: US-08-038-682-2 from: 1 to: 1536
178 AATTAAGGCAAGATTTCAGTCCGGCGCAAGATATTGAGTTTAAACAA 227
199 AsnHisGIyLeuIleThrValGIy...LysAspGIySerVal... 211
228 AAAAGGGAGTGTGGTCCGCAATGCATGACGAAGCCCGATGATGATT 277
212 ....AsnLeuIleGIyGIyLysValLysAsnGIy... 221
278 TTTCTGTGTATCCGCTAACGCG...GTGGCGGCAATTGGCGGCGAT 321
222 ..GIyValIleSerValAsnGIyGIySerIleSerLeuAlaGIyGI 237
322 CA...TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleThrIleAl 254
336 CGTGGCAGATTAACGGCGCTATTAACAATTGATTGGTCCGAGGAA 385
254 AlaIleProGIuAsnGIuAlaValAsnLeuGIyAspIlePheAlaLysGI 271
386 GCAAT...CCGATTCAGCAGCGCTTTCT 411
271 LysIleAsnValAlaArgAlaIleThrIleArgAsnGIyGIyLysLeuSer 287
412 TTCCAATTTGTAAGAATAATTATTAAGCAGGACTAAGCGGCATCC 461
288 AlaAspSerValSerLysAsp...LysSerGIy... 297
462 TTATGCGCGCATATCATATGCGCGTTGCAACAATTTGTCAGAGT 511
298 .....AsnIleValIleuSerA 303
512 CAGAACTGTGAG...ATGACCACTTATATGAT 543
303 LysGIyGIyGIyAlaGIyIleGIyValIleSerAlaGIyAsnGIy 319
544 GGTGGGAATACGCTGATTAAATAATTAACCTGATGCTGT...CG 587
320 GIAlaLysGIyGIyLysLeuMetIleThrGIyAspLysValThrLeuGI 336
588 AATCGAGCA...GGCAGCAATATT 610
336 sThrGIyAlaValIleAspLeuSerGIyLysGIyGIyGIyGIyThrTy 353
611 GCGCGTCTGATGA...GAGCAACCAATAC... 639
353 euGIyGIyAspGIuArgGIyGIyLysAsnGIyIleGIyLeuAlaLys 369
640 .....CGCAAGTTCATATCATATGCAAGC... 666
370 LysThrSerLeuGIyLysGIySerThrIleAsnValSerGIyLysGIy 386
667 .....GCATATTCTTGG...CTGCGGTGGCAATA 694
386 sGIyGIyArgAlaIleValThrGIyAspIleAlaLeuIleAspGIyAsnI 403
695 CTTTGCACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 744
403 leAsnAlaGIy...GIySerGIy... 409
745 AATATTAAACATAGCCCATATGTTTAAACAGAGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGIyGIyPheValGI 419

```

```
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAATGTGT 844
      |||||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGATTTGCAAACAGCAACCCCTATATAGAAAAAGCAAT 894
      |||||
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCACAGTAGTTCGTAAGATTGGTCTATGATGAAATCTTGTCTGG 944
      |||||
436 .....LysGluTrpLeuAsp.....Phe 442
945 AGATACCATTCAGTATCTACAGCACATCAAAATGGAAATACCTTT 994
      |||||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459
995 TTAACGACATATATATAGCGGAGCAAAATCATGCAAAATATAAACAC 1044
      |||||
459 LuAspAspGluIleThrIleSerg.....AsnSer 469
1045 TATCTCTACCTTATAGATTAAACACAGAACCGTTCAATGTTTATGT 1094
      |||||
470 AlaSerThrProLysArgAsnLysGluIleThrThr.....LeuThrAsnThr 485
1095 TTCTTTATCCGAG.....ACAGCA 1114
      |||||
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTAT..... 1128
502 sGlnIleArgIleIleValAsnSerSerIleAsnLeuSerAsnLysSerLeu 518
1129 .....CATGCTGCAAGGTGGGCTCACACTTATTCGACC 1160
      |||||
519 ThrLeuTrpSerGluIleArgSerGlyGlyVal..... 530
1161 CAGACGCAATAT.....GGAGAAATA 1183
      |||||
531 GluIleAsnAsnAspIleThrThrGlyAspSphThrArgGlyAlaAsn 547
1184 TTTCCTTTATTCACAAGAAAAAGTGAATGATGATCTACACGACATC 1233
      |||||
547 eThrIleIleIleThrSergLysGlyIleThrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGGCGGCTTTGATTTTGAGGGTAAATTTACGCTCTCGCC 1283
      |||||
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAla 573
1284 TAA.....AACACGAACGCTGCGCAGCGC 1309
      |||||
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGly 590
1310 CGGCGCTTCATATCAGTATGATGACGATACCGTTACTTGAAAGTAAACGC 1359
      |||||
590 InGlyThr.....IleThrSergLysAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAACACCGCTCTCCAAATCGCAAAAGCGACGCTGCTGATCA 1409
      |||||
606 ValSer.....LeuAsnGlyThrGlySerg.....LeuGlnPheThr 618
1410 AGCCAAAGGGGAAAC.....CAAGGCTCG 1435
      |||||
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluIleThr 635
1436 TCACGCTGGCGGCGGTAAGTCACTTA.....GAT 1467
      |||||
635 euAsnIle.....SergLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGGCGGATCAAGCAAAACAAAGCCTT..... 1503
      |||||
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....AGTGAATCGGC.....TTGGTCAGCGGCGAGG 1531
```

```
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArg 684
      |||||
1532 GCAGC.....GNG 1539
      |||||
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACGTGAATGCCGATATCATGTTCAACCCGACAAA..... 1575
      |||||
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsn 734
1580 ATTTCGGCTTCGCGCGGACGCTTGATTTGAACGGCATCGCTTTCG 1629
      |||||
734 rAlaSerPheAsnGlyAsnIleSerValSergLysGlySerValAsp 750
1630 TTCACCGCAT.....CAAAATCCGATGAAGGGCGATGATGT 1670
      |||||
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValAla 767
1671 CAACCAAT.....CAGACCAAG 1690
      |||||
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLys 784
1691 AATCCACCGTTACCATTCAGCG.....ATAAAGATATTA 1728
      |||||
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 .....ACAAACCGCAATAAC..... 1743
      |||||
801 AsnAlaThrArgGlyAlaAsnIleThrLeuLeuGlnValGluGlyThrAsp 817
1744 .....AACACTTGATACCAAAAGAAATGCCACACAGGT 1782
      |||||
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGly 834
1783 .....TGCTTGGCGGCAAGATGCA...ACCAAAACGACGAGCGG 1821
      |||||
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluLys 850
1822 CTCATTCGATTTACCAACCGGACGACGATCGCACTTA..... 1863
      |||||
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881
1864 .....CTGCTTCCGCGCGA...ACAAATTTAAACGC 1893
      |||||
881 leAsnSerGlyAsnLeuThrAlaGlyLysAlaIleValAsnIleAlaGly 897
1894 AATATCAGC...CAACAAACGCGCAACTGTTTTCACGCGCA..... 1933
      |||||
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934 .....GACCGACCGCGCGCTCAATC 1957
      |||||
913 heThrPheAsnValGlyIleLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTAGAAGCGGTGTCAAAATGAGATATCCACAGAGAAATC 2007
      |||||
930 SerIleAlaLysGlyLysAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTCTGGACACAGCATTCATCGACCGC.ACATTTAAAGCGGAAACTTCC 2056
      |||||
946 nLeuSerIleThrThrAsnSerSerSerThrTyrArg..... 958
2057 ATATTGAGGCGGACAAAGCGGTGCTTCCCGCAATGTTCCAAAGTGAA 2106
      |||||
```

```

959 .....ThrlleleserGlyasnlethrAsnLysasn 969
2107 GGCATTGGCATTTAGCAATTCACGCC.....CAACGACTTTT 2144
||||| :||||| :||||| :||||| :||||| :|||||
970 GlyAspLysasnlethrAsnGlySerAspThrGluIleThrIleG 986
2145 CGGTGTCGACCGCCATCAAGCCACACATCTGTACAGCTTGCGATGA 2194
||||| :||||| :||||| :||||| :||||| :|||||
986 YGlyAspValSerGlnGlyGlnGlyAsnlethrIleSerSerAspLysI 1003
2195 CGGCTGACAGAGTTTACCGAATAAACCTTACC...GACGATAAAGTG 2241
||||| :||||| :||||| :||||| :||||| :|||||
1003 Leasnlethr.....LysGlnlethrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGAGCAAGACGACATCAGACGATGTCAGCTTGCCGA 2291
||||| :||||| :||||| :||||| :||||| :|||||
1016 AspGlyGlnAsnSerAspSerAspAlaThrAsnAlaAsnlethrI 1032
2292 TCACGCT.....CATTTAAATCTCACAGAGACTT. 2319
||||| :||||| :||||| :||||| :||||| :|||||
1032 eLysThrLysGluLeuLysleuThrGlnAspLeuAsnleSerGlyPhe 1049
2320 .....GCCACTCAACGCC 2334
||||| :||||| :||||| :||||| :||||| :|||||
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlysn 1065
2335 AATCTTAGTCAGCGCGAGACACGCACTATACGCTTACGCCACGCCAC 2384
||||| :||||| :||||| :||||| :||||| :|||||
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysValThrPheAsnGlnI 1082
2385 CCAAAACGCAACCTCAGCTCGTGCGCAATGCCACAGCAATTATTC 2434
||||| :||||| :||||| :||||| :||||| :|||||
1082 LysAspSerLysIleSerAlaAspGlyHisValThrLeuHisSerL 1099
2435 AAGCCATTAATTAACGCGACACATCGCTTCGACATGCTTCATTAT 2484
||||| :||||| :||||| :||||| :||||| :|||||
1099 YsValGluThrSerGlySerAsnAsnThrGlnAspSerSerAspAsn 1115
2485 CTAAAGCAACAGCGGTACAAACAGCGAGTGCAGCTTCCGCAACGC 2534
||||| :||||| :||||| :||||| :||||| :|||||
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValaAsnAsn 1132
2535 TAAGGCAAC.....GTAAGCATTCGCGACTCAACGCAATGTC... 2574
||||| :||||| :||||| :||||| :||||| :|||||
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyIleThr 1149
2575 .....TCCTAGCCGTAAGGCAATTCATTGCAAAACAGC 2613
||||| :||||| :||||| :||||| :||||| :|||||
1149 hLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIle 1165
2614 CGCTTACCGGAAATCAGCGCGGCAAGATAG.....GCATT 2654
||||| :||||| :||||| :||||| :||||| :|||||
1166 AlaGlnThrGlySerIleLeuGlyGlyIleLysSerSerGlySerVa 1182
2655 ACACTTAAAGACAGCGAATGACGCTCGCTCGGACAGGAATTAGCA 2704
||||| :||||| :||||| :||||| :||||| :|||||
1182 IThrleuThrAlaThrGlnGlyAlaLeu.....AlaValSerA 1195
2705 ATTAAACCTTGACACAGCGCATTTACATCACTTCGCCCTATCGACAC 2754
||||| :||||| :||||| :||||| :||||| :|||||
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGCGAGCGCGCAACCGGACGATGACGCGCGCGCGCGCG 2804
||||| :||||| :||||| :||||| :||||| :|||||
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
2805 TTCGCGCGCTTCCTATTATTCCTTACGCGCACTTCGCGAAGATCC 2854
||||| :||||| :||||| :||||| :||||| :|||||
1217 .ThrIleGlyGlyThrGlnSerValThr.....ThrSerGlnSerG 1231
2855 GTTTCACAGCGTGAAGGTAAGGCAATTGACGTAAGGACACA... 2901
||||| :||||| :||||| :||||| :||||| :|||||
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242

```

```

seq.name: /cgn2_6/ptodata/1/laa/5A_COMB.pep:us-08-302-832-2

```

```

seq.documentation_block:

```

```

: Sequence 2, Application US/08302832

```

```

: Patent No. 5603938

```

```

: GENERAL INFORMATION:

```

```

: APPLICANT: Barenkamp, Stephen J

```

```

: TITLE OF INVENTION: High Molecular Weight Surface Proteins

```

```

: NUMBER OF SEQUENCES: 8

```

```

: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Shoemaker and Mattare, Ltd.

```

```

: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

```

```

: STREET: Bldg. 1

```

```

: CITY: Arlington

```

```

: STATE: Virginia

```

```

: COUNTRY: U.S.A.

```

```

: ZIP: 22202-0286

```

```

: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk

```

```

: OPERATING SYSTEM: IBM PC compatible

```

```

: SOFTWARE: Patent In Release #1.0, Version #1.30

```

```

: CURRENT APPLICATION DATA:

```

```

: APPLICATION NUMBER: US/08/302,832

```

```

: FILING DATE: 16-SEP-1994

```

```

: CLASSIFICATION: 435

```

```

: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: GB 9205704.1

```

```

: FILING DATE: 16-MAR-1992

```

```

: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US pct/us93/02166

```

```

: FILING DATE: 16-MAR-1993

```

```

: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Berkstreser, Jerry W

```

```

: REGISTRATION NUMBER: 22,651

```

```

: REFERENCE/DOCKET NUMBER: 1038-404

```

```

: TELEPHONE: (703) 415-0810

```

```

: TELEFAX: (703) 415-0813

```

```

: INFORMATION FOR SEQ ID NO: 2:

```

```

: SEQUENCE CHARACTERISTICS:

```

```

: LENGTH: 1536 amino acids

```

```

: TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

```

	alignment_scores:	
quality:	265.00	length: 1297
ratio:	0.438	gaps: 73
Percent Similarity:	46.646	Percent Identity: 21.126

alignment_block: MS-08-302-832-2
MS-08-303-518D-653

Align seg 1/1 to: US-08-302-832-2 from: 1 to: 1536

```

178 AATMAAGCGCAAGTTTCGACGTGGGGCGCAAGATATTGAGTTTTCACAA 227
      ||| ||| :|||:||||||| ||||| |||||
199 AsnHisIleuLeuIleThrValGly..LysAspIleSerVal..... 211
      :|||:||||| ||| :|||: |||
212 .....AsnLeuIleIleGlyLysValLysAsnLu..... 221
      :|||:||||| ||| :|||: |||
278 TTTCTGTGATTCGCGGTACGGC.....GTGGCGCATTTGGCGGGCAT 321
      :|||:||||| ||||| :|||: |||||:|||||
222 ..GlyValIleSerValAsnGlyIleSerIleSerLeuLeuAlaGlyGln 237
      :|||:||||| ||| :|||: |||
322 CAA.....TTATTGTGGAG 335
      :|||: |||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
      :|||: |||
336 CGTGGCAATTAACGGCGGTATTAACAATGTGATTTTGTCGGCAGGAA 385
      :|||: ||| ||| :|||: |||:|||||
254 AlaProGluAsnGlnAlaValAsnLeuGlyAspIlePheAlaLysGly 271
      :|||: |||
386 GCAT.....CCGATCAGCACCGCTTCT 411
      :|||: |||:||||| :|||: |||
271 LysAlaIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287
      :|||: |||:||||| |||: |||
412 TACCAAAATTTGAAAGAAATATTATTAAGCAGGCACTAACGCCATCC 461
      :|||: |||:||||| |||: |||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
      :|||: |||
462 TTATGGCGGCATTTATCATATGCCGCTTTCACACAATTTGTCAAGANG 511
      :|||: ||| :|||: |||
298 .....AsnIleValLeuSerA 303
      :|||: |||
512 CAGAACCTGTGGAG.....ATGACCGATTATATGTGAT 543
      :|||: ||| :|||: |||
303 IAlaValGluGlyGlnAlaGluIleGlyGlyValIleSerLAlaGlnAsnGln 319
      :|||: |||
544 GGGTGGAAATACGCTGATTTAAATAAATACCTGATCGTGT.....CG 587
      :|||: ||| ||| :|||: |||
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuVal 336
      :|||: |||
588 AATCGGAGA.....GGCAGCAATATT 610
      :|||: ||| ||| :|||: |||
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
      :|||: |||
611 GCGCGTGTATAA..GAGCAACCAATTAAC..... 639
      :|||: ||| ||| :|||: |||
353 euGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLys 369
      :|||: |||
640 .....CGCGAAATCATATCATATTCGAAC..... 666
      :|||: |||:||||| :|||: |||
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlu 386
      :|||: |||
667 .....GCATATTCTTG.....CTCGTCGGTGGCAATA 694
      :|||: ||| ||| :|||: |||
386 sGlyGlyAlaAlaIleValThrPheLysPheAlaLeuIleAspLysAsnI 403
      :|||: |||
695 CCTTTGCACAAATGATTCAGGTGGTGGCACAGTCAACTTAGTAGGCA 744

```

```

403 leasnlaagln...glysergly..... 409
745 AAAATTAACATAGCCCAATATGCTTTTTCACAAAGAGGCTCATTTGG 794
410 .....AspIlealalysthrGlyIypheValGl 419
795 CGAAGAGTGCTCAACCAATGTTATCTATGATGCGCCAAAGCAAAAGTGAT 844
419 urthSerGlyHisAspIeuhellelyasp... 429
845 TAATTAATGGGATATTGCAACACGCAACCCCTATATAGAAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCCTCCAGCTACTTGCTAAAGATGGTGTCTATGATGAATCTTTGCTGG 944
436 .....LysGluTPLeuAsp.....Phe 442
945 AGATACCCATTCAGTATTTCTACGAAACACATCAAAATGGCAATACTTTT 994
442 eaSpasnValSerIleasnIaGluThrIaGluArgSerAsnThrSerg 459
995 TTAAAGCAATATATATAGCGCGCCAGGAAAAATCGATGCCAAACATAAAC 1044
459 lnaSpasglutylThrGlySergly.....AsnSer 469
1045 TATTCCTACCTTATAGATTAAAAACGCAACCGCTTCATTTGTTAATGT 1094
470 AlaSerThrProLysArgAsnLysGluIysThr...LeuThrAsn 485
1095 TTTCTTTATCCGAG.....ACAGCA 1114
485 rThrIeuGluSerIleleuLysGlyThrPheValasnIleThrAla 502
1115 GAGAACCTTTAT..... 1128
502 snGlnArgIleLysValAsnSerSerIleasnLeuSerAsnGlySerLeu 518
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVala..... 530
1129 .....CATGCTGCAAGTGCGGTCAACAGTTATTCGACC 1160
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVala..... 530
1161 CAGACTGAATAT.....GGAGAAATA 1183
531 GlnIleAsnAsnAspIleThrThrGlyAspAspThrArgIlyAlaAsn 547
1184 TTTCTTTTGTTCGAAAGCAAAAGGTGATGATCTTACACAGACATC 1233
547 eutThrIleLysSerGlyGlyTPLeuAsp.....ValHisIlyAsnIle 561
1234 AACCAAGCGCGCGCGGTGTGATTTTGAAGGTATTTTACAGCTCGCC 1283
562 SerLeuGlyAla.....GlnIlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACCAAGCAAGTGCACAAAGCGC 1309
573 alySGlnAspIleAlaPheGluLysGlySerIleAsnIleValIleThrGly 590
1310 CGGCGCTTCATTAACAGTATGCGAGTACCGTTACTTGGAAAGTAAAGCG 1359
590 lnglyThr...IleThrSerGlyAsnGluLysGlyPheArgPheAsnAsn 605
1360 GTGGCAACAGCAGCGCTGTGTCAAAAATCGGCAAGGCAAGCTCTGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySergly...LeuGlnPheThr 618
1410 AGCCAAAGGGGAAAC.....CAAGCTCGC 1435
618 rThrLysArgThrAsnLysArgAlaIleThrAsnLysPheGluGlyThrL 635
1436 TCACGCTGGCGAGCGTAAAGTATCTTA.....GAT 1467

```

635 euansnle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CACGAGCGGCGATCAAGCAGCAAAAAACAGCGTTT... 1503
651 GluSerGlyTyrAspLysPheLysGlyAlaGlyThrTyrTrpAsnLeuThrSe 667
1504ACTGAATCGGC.....TTGGTCAGCGCGCAGG 1531
667 rLeuAsnValSerGluSerGlyLupheAsnLeuThrIleAspSerArg 684
1532 GGAGC.....GTG 1539
684 LysSerSpSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyLe 700
1540 CAACGTAAATGCGGATATCATGATTCACCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsn 734
1580 ATTTCGGCTTTGCGCGGACGCTTGTGAACGGCGCATTCGCTTTCG 1629
734 yAlaSerPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACACCGCAT.....CAAAATACGATGAGGGCGATGATGT 1670
751 PheThrLeuLeuAlaSerSerAsnValGlnThrProGlyValIle 767
1671 CACCCACAT.....CACACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuThrPheLys 784
1691 AATCCACCGTTACCATTAACGCG.....AATAAGATATTACT... 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGlyLysAspLeuThrLeu 800
1729 ...ACACCCGCAATTAAC..... 1743
801 AsnAlaThrGlyLysIleThrLeuLeuGlnValGlnLysThrAspG 817
1744AACAACCTTGATACCAAAAAAATAATGCTTAACAGCT 1782
817 yMetIleGlySerGlyLeuAlaLysLysAsnIleThrPheGlnGlyG 834
1783TGCTTGGCGAGAAAGATGCA...ACCAAAACGAGGCGCG 1821
834 LysAsnIleThrPheGlySerArgLysAlaValThrGlnIleGlnLys 850
1822 CTCATCTGATTAATCAACCGCAAGGCGGATCGACATTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881
1864CTGCTTCCGCGGCA.....ACAATTTAAACGCG 1893
881 LeuAsnSerGlyAsnLeuThrAlaGlyLysIleValAsnIleAlaGly 897
1894 AATATACAG...CAACAAACGCGCAATGTTTTCAGCGGCA..... 1933
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934GACCGACACCGCAGCGCTCAATC 1957
913 heThrPheAsnValGlyLysLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTCAGACGCGGTGTCAAAATGAGAGATATGCCACAGAGAAATC 2007
930 SerIleAlaLysGlyLysAlaThrPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGAGACAGCATTTGATCGACGCG..ACATTTAAACGGAACCTCC 2056
946 nLeuSerIleThrThrAsnSerSerThrTyrArg..... 958
2057 ATATTACGGCGGACACAGGTGGTTTCCCGCATGTTGCCAAAGTGGA 2106
959ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGGATTCGATTTAAGCAATCAGCG.....CAAGAGTTT 2144
970 GlyAspLeuAsnIleThrAsnGlnLysSerAspThrGlnMetGlnIleG 986
2145 CGGTGCGGACCGCATCAAGGCAACACATGTGACATCGCTGCGATGGA 2194
986 yGlyAspValSerGlnLysGlnLysLeuThrIleSerSerAspLysI 1003
2195 CGGTCTGACAGTTGTACGAAAAAACCATTACC...GACGATAAATG 2241
1003 LeuAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGACGACAGCAGCATCAGAGCAATGTCAGCTTGCGGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrI 1032
2292 TCACGT.....CATTTAATCTCAGAGACTT 2319
1032 eLysThrLysGlnLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049
2320GCCACGTCACAGCG 2334
1049 snLysAlaGlnIleThrAlaLysAspLysSerAspLeuThrIleGlnLys 1065
2335 AATCTTATGCGAGCGGACAGACGCACTAATACGTTACGCCGACCGCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVal 1082
2385 CCAAAACGCGCAACCTGACCTCGGCGCAATGCCCAAGCAACTTAATC 2434
1082 LysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSer 1099
2435 AAGCCACATTAAACGCGACACATCGGCTTCGACATGCTTCATTAAAT 2484
1099 yValGlnThrSerGlySerAsnAsnAsnThrGlnLysPheSerSerAsp 1115
2485 CTAGCAACAAGCGCGTACAAAGGCGAGCTGACGCTTCCGACAAACG 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsnI 1132
2535 TAAGCAAAAC.....GTAAAGCATTCGCGACATCAAGGCAATGTC 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGlnIleThr 1149
2575TCCCTAAGCCGATTAAGGCAATTCGATTTGAAAAACAG 2613
1149 hrLysThrGlyThrThrIleAsnAlaIleThrGlyAsnValGlnIleThr 1165
2614 CGCTTAACGGAATAATCAAGCGCGGCAAGGATAG.....GCAT 2654
1166 AlaGlnThrGlySerIleLeuGlnLysIleGluSerSerGlySerVal 1182
2655 ACATTTAAAAAGACAGGATGAGCGCTGCGGCGACAGGATTAAGCA 2704
1182 ThrLeuThrAlaIleThrGlnLysAlaLeu.....AlaValSerA 1195
2705 ATTTAAACCTTGACAAAGCGACATTAACATCAATTCGCTATCGACAC 2754
1195 snLysSerLysAsnThrValThrAlaAsnSer..... 1207
2755 GATGCGGCGGCGCGCAACCGCGAGTCGCGACAGATGCGCGCGCGCG 2804
1208GlyAlaLeuThrThrLeuAlaGlySer..... 1216


```

2805 TTGCGCCGCTTCCCTTATATCCGTACGCCGCAACTTCCGAGCAATCC 2854
      ::::: ||||| ||||| ||||| |||||
1217 .ThrlleYgLyThrgluserValthr.....ThrserserInserc 1231
2855 GTTTCACACGCTGACGTAACGCAAAATTGACGCTGACGAGACA... 2901
      ::::: ||||| ||||| ||||| |||||
1231 ly.....AsplleYgLyThrlleSerserly...GlyThrlVal 1242
2902 ...TTCCGCTTATGTCGAACTCTGCTACCGAGCGGCAAAATTGAA 2948
      ::::: ||||| ||||| ||||| |||||
1243 GluVallysalathrgluserleuthrThrglInserthsnserlylely 1259
2949 GCTGGCGGAAAGTTCGAGGCACTTACACCTTGCTGTC..... 2988
      ||| ||| ::::: ||||| ||||| |||||
1259 s...AlathrThrglYgluAlasnValThrsAlathrglyThrlleG 1275
2989 .....ACAAATTCGCGGCAACGACCCGTAATCTCGAGCAA 3024
      ::||| ||| ||| ::::: ||||| |||||
1275 lyglYThrlleSerserlyasnThrValasnValThrlAlasnaAlaglyasp 1291
3025 TTGACGGTA.....GTGAGAGAAAGACAA 3050
      ||||| ||| ::||| ||| ::||| |||
1292 leuthrValIglYasnGlyAlagluIleasnAlathrglYgluAlaIath 1308
3051 CACACCGCTGTCGAAATTTATTCACCTGCAAAACGACACGCTCG 3100
      ::||| ||| ||| ::||| ||| |||
1308 rleuthrThrserserlyLysleuthrThrglualaserSerthslter 1325
3101 ATCGCGGGGCAATGGGCTATACGCTTACGCAAGACGGGAGATTCGC 3150
      ::||| ||| ||| ::||| ||| |||
1325 hrserAlalysGlygluValasnleuserAlaglInspglySerValAl 1341
3151 CTGCATATTCGCTCAAGAACAGACCTTCCGACA 3187
      ||| ||| ::||| ||| ::||| |||
1341 aclySerleleasnAlaAlasnValThrlleuasnThr 1353
seq_name: /cgn2_6/ptodata/1/1aa/5b_COMB.pep:US-08-530-198-2

```

```

seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

;
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-2

```

```

alignment_scores:
  Quality: 265.00      Length: 1297
  Ratio: 0.438        Gaps: 73
  Percent Similarity: 46.646      Percent Identity: 21.126

```

alignment_block:

US-09-303-518D-653 x US-08-530-198-2

Align seg 1/1 to: US-08-530-198-2 from: 1 to: 1536

```

178 AATAAGGCAAGTTTGCAGTCGGGCGAAGATATTGAGGTTTACAA 227
    ||| ||| ::||| ||||| ||||| |||||
199 AsnHsGlyleuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAGGGAGTTGGTCGGCAATCGATGACGAAAGCCCGATCATGATT 277
    ::||| ||| ||| ::||| |||
212 .....AsnleuIleGlyLysValLysAsnGlu..... 221
278 TTTCGTGTATCGCTACGCG.....GTGGCGCATTCGGCGCGAT 321
    ::||| ||| ||| ||||| |||||
222 .GlyValIleSerValasnGlySerIleSerleuIleAlaglyGln 237
322 GAA.....TATATTGAG 335
    ::||| |||
238 LysIleThrlleSerAspIleIleasnProThrlleThrlySerIleAl 254
336 CCGTGCATATACGGCGGCTATACCAATGTGATTGGTCGGAGGCA 385
    ::||| ||| ::||| ||||| |||||
254 alaProGluasnGluAlaValasnleuGlyAspIleIleAlaIysGly 271
386 GCAAT.....CCGATCAGCACCGCTTTTCT 411
    ::||| |||
271 LysAsnIleasnValargAlaIathrlleargasnGlnGlyLeuser 287
412 TACCAATTTGCAAGAATATATTAATTAACGAGGACTACGCGCATCC 461
    ::||| ||| ||| ::||| |||
288 AlaAspSerValSerLysasp.....LysSerGly..... 297
462 TTATGCGCGCATATATCATATCCGCGTTGCACAAATTTGTCACAGATG 511
    ::||| |||
298 .....AsnIleValleuserA 303
512 CAGAACCTGTGAG.....ATGACCAATTATATGAT 543
    ||| |||
303 lalysgluGlygluAlagluIleGlyValIleSerAlaGlnasnGln 319
544 GGGTGCAAATACCGTATTAATAATACCTGATCGTGT.....CG 587
    ||| ||| ||| ||||| |||||
320 GlnAlaIysGlyLysLysleuMetIleThrglyAspLysValThrllely 336
588 AATCGAGACA.....GGCAGACAAATATT 610
    ||||| |||||
336 stInrglyAlaValIleAspLeuserGlyLysgluGlygluThrlyTL 353
611 GCGGCTGTGATGA...GACGAACCAATAC..... 639
    ::||| ||| ||| ::||| |||
353 euGlyglYaspGluargGlygluGlyLysasnGlyIleGlnleuAlaIys 369
640 .....CGCGAAGTTCATATCATATTGCAAGC..... 666
    ::||| ||| ||| ::||| |||
370 LysThrsleuGluLysGlySerThrlleasnValserGlyLysgluLy 386
667 .....GCATATTTCTTG.....CTCGTCGGTGGCAATA 694
    ||| ||| |||
386 sGlyglYargAlaIleValThrpolYaspIleAlaIleAspGlyAsnI 403

```

```

695 CCTTGCACAAAATGATCAGGTGGTGGACAGTCACTTAGGTAGCGAA 744
    |||||
403 leasnaIaGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGTTTTCACCAACAGAGGCTCATTTGG 794
    |||||
410 .....AspIleAlaIysThrGlyGlyIleValG1 419
795 CGACAGTGGCTCACCAGTGTTCATTCATGATGCCAAAAGCAAAATGGT 844
    |||||
419 uThrSerGlyHisAspIeuPheIleIysAsp..... 429
845 TAAATTAATGGGTATTCGCAACAGCAACCCCTATATAGAAAAAGCAAT 894
    |||||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTCCAGCTAGTTCGTAAGATGTGTTCTATGATGAATCTTGGCTGG 944
    |||||
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTCACAGCACATCAAAATGGCAATACCTTT 994
    |||||
442 eaSpAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459
995 TTAAACGCAATATATATGGCGCAGGAAAAATCGATGCCAAACATAAGAC 1044
    |||||
459 IuAspAspGluTrpThrGlySerGly.....AsnSer 469
1045 TATTCCTCAGCTATATAGATTAAACAGCAACCGTTCATGTTTAATGT 1094
    |||||
470 AlaSerThrProLysArgAsnLysGluYsnThr...LeuThrAsnThr 485
1095 TTCTTTATCCGAG.....ACAGCA 1114
    |||||
485 rThrIeuGluSerIleLeuLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTAT..... 1128
    |||||
502 snGlnArgIleThrValAsnSerSerIleAsnLeuSerAsnGlySerIeu 518
1129 .....CATGCTCAGAGTGGGTCAACACTATATCCAGCC 1160
    |||||
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVal..... 530
1161 CAGACGGAATAT.....CGAGAAATA 1183
    |||||
531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTGCAAGAAAGTGAATGTGATACCTTACCAAGATC 1233
    |||||
547 euThrIleThrSerGlyGlyTrpValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGGCTTGTATTTGAGGTATTTACGCTCCGCC 1283
    |||||
562 SerIeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACACGGAACGTGGCAGCGC 1309
    |||||
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGCGCTTCATATCAGTATGAGTACCGTTCATTCGAAAGTAAACGGC 1359
    |||||
590 InGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGCAAAACGACCGCTGCCAAAATCGCAAAAGCAACGCTGCTGTTC 1409
    |||||
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGGAAAAAC.....CAAGCTCGG 1435
    |||||
618 rThrLysArgThrAsnLysThrAlaIleThrAsnLysPheGluGlyThrL 635

```

```

1436 TCACGTGGCGCGAGGTAAAGTCATCTTA.....GAT 1467
    |||||
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGCGAGTCACAGGCAAAAACAAACCTTT..... 1503
    |||||
651 GluSerGlyTrpAspLysPheLysGlyArgThrTrpAsnLeuThrSe 667
1504 .....ArgMAATCGC.....TGGTCAGCGGAGCG 1531
    |||||
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArg 684
1532 GAGCG.....GTG 1539
    |||||
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyAsnLeuAsnGlyLe 700
1540 CAACGTGAATCCGATATCATGTTCAACCCCGACAA..... 1575
    |||||
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgVal 717
1576 .....CCT 1579
717 nPheAspIleLysAlaProIleGlyLysAsnLysTrpSerSerIeuAnt 734
1580 ATTTGGCTTTCGGCGGCGAGCTTTGATTTGAACGGCATTGCTTTCG 1629
    |||||
734 YrLaserPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGGATT.....CAAAATACGATGAAGGGGAGGATGTTGT 1670
    |||||
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValIle 767
1671 CAACCAAT.....CAAGCAAG 1690
    |||||
767 eaSnSerLysTrpPheAsnValSerThrGlySerSerLeuArgPheLys 784
1691 AATCCACCGTTTACCATTCACAGC.....AATAAGATATTACT... 1728
    |||||
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 ...ACACCGGCATATAC..... 1743
    |||||
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAsp 817
1744 .....AACAACTTGATACCAAAAAGAAATGCTACACAGCT... 1782
    |||||
817 YmetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGly 834
1783 .....TGGTTGGCGAGAAAGATGCA...ACCAAAACGACGCGCG 1821
    |||||
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluLys 850
1822 CTCATATCGAATTACCAACCGGAAGAGCGGATCGCACTTAA..... 1863
    |||||
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisLysLysProLeuThrIleLysLysAspValIle 881
1864 .....CTGCTTTCGCGCGA.....ACAAATTAAACGCG 1893
    |||||
881 leAsnSerGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGly 897
1894 AATATCAG...CAAAACAAAGCGCAAACTTTTTCAGCGCA..... 1933
    |||||
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsp 913
1934 .....GACCGACCGCAGCGCTCAATC 1957
    |||||
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTAGGAACGGGTGTCAAAATGGAAGATATCCACACAGGAGAAATC 2007

```

```

930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGGGACAAACGATGGATCGACCGC..ACATTTAAAGCGGAAACTTCC 2056
946 nLeuSerIleThrAsnSerSerSerThrTyrArg..... 958
2057 ATATTCAAGCGGACAAAGCGGTGGTTCCCGCAATGTTCGCAAAAGTGAA 2106
959 .....ThrIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGCATTTGACATTACGATCAGCC.....CAGACAGTTT 2144
970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGluIleG 986
2145 CGGTGTCGACCGCATCAAGCCACACATCTGTACAGTTCGCGACTGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003
2195 CGGCTGTACAGTTGTACGAAACCATTACC...GACGATAAAGTG 2241
1003 LeAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTATGTACGACAGACGACATGACGCAATGTACGCTTCCGCA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrI 1032
2292 TCACGCT.....CATTTAAATCTCACAGGACTT. 2319
1032 eLysThrLysGluLeuLysLeuThrGluAsnLeuIleSerGlyPheA 1049
2320 .....GCCACACTCAAGCGC 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTCGCGGAGACGACGACTATACGTTAGCGCGCAAGCGC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGluVa 1082
2385 CCAAAACGGCAACTCAGCCTGTGGGCAATGCCACAGCAATTTAATC 2434
1082 LysAspSerLysIleSerAlaAspGlyHisValThrLeuHisSerL 1099
2435 AAGCCCATTTAAAGCGCAACATCGGCTTGCACATCTTCATTAT 2484
1099 yValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAAGCAACAGCGCTGACAAAACGCGACTGACGCTTCCGACACGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnI 1132
2535 TAAAGCAAC.....GTAAGCATTCGCGACACAGCGCATGTC.... 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149
2575 .....TCCCTAGCCGATTAAGCAGATATTCATTTTGAACACGC 2613
1149 hTrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165
2614 CGCTTACCGGAAATAATCAGCGGCGCAAGGATAGC.....GCATT 2654
1166 AlagLThrLysSerIleLeuGlyLysIleGluSerSerSerGlySerVa 1182
2655 ACACCTTAAAGACAGCAATGACGCTGCGGTGGGCGACGATTTAGGA 2704
1182 lThrLeuThrAlaThrGluGlyAlaLeu.....AlaValSerA 1195
2705 ATTTAAACCTTGACACGCGACCATTAACCTAATTCGCTATCGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGGCGAGCGCGCAACCGGCAAGTGCGGACAGATCGCGCGCGCG 2804

```

```

1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
2805 TTGCGCGCCTTCCCTATTATTCGTTACGCGCGCACTTCGCGAATGCC 2854
1217 .ThrIleLysGlyThrGluSerValThr.....ThrSerSerGlnSerG 1231
2855 GTTTCAACACGCTGACGTTAAAGCGCAATGTACGCTCAGGACAA... 2901
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
2902 ....TTCCGCTTATGTGCGAAGCTTTCGCTACCGCAGCGCAATTTGA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIle 1259
2949 GCTGCGGGAAGTTCGGAAGCGCACTTACACCTTGGCTGTC..... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275
2989 .....AACATATCCGCGCAACGAAACCGCTAAGTCTCGAGCA 3024
1275 LysGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291
3025 TTGACGCTA.....GTGGAAGGAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThr 1308
3051 CACACCGCTGCGCAAAATCTTAATTTCACCTGCAAAACGACAGCTCG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIle 1325
3101 ATGCGGCGGATGGGCTTATACGCTTATCGGCAAGACGCGAGTTCCGC 3150
1325 hrSerAlaLysGlyGluValAsnLeuSerAlaGluAspLysSerValAl 1341
3151 CTGCATATCCGTCGCAAGAAACAGAGCTTTCGACA 3187
1341 aGlySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

```

seq_name: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-469-880-2

```

seq_documentation block:
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832

```

```

?      FILING DATE: 16-SEP-1994
?      ATTORNEY/AGENT INFORMATION:
?      NAME:  Berkstresser, Jerry W
?      REGISTRATION NUMBER: 22,651
?      REFERENCE/DOCKET NUMBER: 1038-516 MTS-VG
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (703) 415-0810
?      TELEFAX: (703) 415-0813
?      INFORMATION FOR SEQ ID NO: 2:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 1536 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      OS-08-469-880-2

```

alignment_scores:		
Quality:	265.00	Length: 1297
Ratio:	0.438	Gaps: 73
Percent Similarity:	46.646	Percent Identity: 21.126

alignment_block:

US-09-303-518D-653 X US-08-469-880-2 .

Align seg 1/1 to: US-08-469-880-2 from: 1 to: 1536

[illegible][illegible]

590 lnglythr...llethrserglyasnglulysglypheargpheasn 605
 1360 GTGGCAACGACGCGCTGTCACAAATGCGCAAGGACGCTGCTGTCA 1409
 606 valser...leuasnnglythrlysergly...leuglnpnerh 618
 1410 AGCCAAAGGCGGAAAC...CAAGGCTCGG 1435
 618 rthrlyargthrlnslystyralailethrlnslyphegluglythr 635
 1436 TCACGCGGGGCGAGGTAAAGTCATCTTA...GAT 1467
 635 euannile...serglyvalasnleisermetvalleuprolysasn 650
 1468 CAGCAGGCGGATCAAGCAAAACCAAGCCTTT... 1503
 651 gluserglytyrasplysphenylglyargthrytrpansleuthrse 667
 1504 ...AGTAATCGGC...TTGGTCAGCGGCGAGG 1531
 667 rleuasnvalsergluserglyglupheasnleuthrileaspserrg 684
 1532 GGAGC...CTG 1539
 684 lyseraspserralaglythrleuthrlnprotyrasnleuasnlyle 700
 1540 CAACTGATGCGGATATCAAGTTCACCCGACAAA... 1575
 701 serpheasnlyasprthrthrphasnvalgluargasnalaargvalas 717
 1576 ...CTCT 1579
 717 nphasnleuysalaproileglyleasnlystyserleuasn 734
 1580 ATTGGGCTTTGGCGGCGAGCTTTGATTCAGCGCATTCGCTTCG 1629
 734 yralserpheasnlyasnleiservalserglyglyseralasp 750
 1630 TTCACCGCGATT...CAAAATACGATGAAGGGCGATATTGT 1670
 751 phethrleuvalaserseaserasnvalglthrproglyvala 767
 1671 CAACCAAT...CAAGCAAG 1690
 767 easnserlytyrphasnvalserthrlyserleuargpelyst 784
 1691 AATCCACCGTTACCATTCAGGC...AATTAAGATATTACT... 1728
 784 hserglyserthrlysthnglypheserleuglulysaspleuthrleu 800
 1729 ...ACAACGCGCAATAC... 1743
 801 asnalathrglyglasnlethrleuvalaiglylythraspel 817
 1744 ...ANCACTGGATGCAAAAGAAATTCGCTACACGCT... 1782
 817 ymetlileglylysglylevalalalylysasnlethrphlegllyg 834
 1783 ...TGTTGGCGAGAAAGATGCA...ACCAAAAGCAAGGCGG 1821
 834 lysnlethrphleglyserarglysalvalthrlnleuglyasn 850
 1822 CTCATCTGATTAACCAACGGAAGAGCGGATCGCATCTTA... 1863
 851 valthrleasn...asnasnalaasnvalthrleuileglyse 864
 1863 ... 1863
 864 raspphaspanshisglulysproleuthrlelylyaspyalile 881
 1864 ...CTGCTTCCGCGGGA...ACAATTTAAACGC 1893
 881 leasnserglyasnleuthrthralaglyglasnlevalasnleleagly 897

1894 AATATCAGC...CAAAACAGCGCAACGTGTTTCAGCGCA... 1933
 898 asnleuthrvalgluserasn...phenylsalalethrnsrp 913
 1934 ...GACCGACACCGGACGCTTCAATC 1957
 913 hethrphasnvalaiglyleupheasnlygllyasnserasnle 929
 1958 ATTAGGAAGCGGTGTCAAAAATGGAAGATATCCACACGACAAATC 2007
 930 serllealalysglylalaargpelysaspilleaspsnserlyasn 946
 2008 GTCGTGGCAACGATGATCGACGC...ACATTTAAAGCGGAAACTTCC 2056
 946 nleuserlethrthrasnserseerthrlyarg... 958
 2057 ATATTCAGGCGGACAGCGGTTGTTCCGCATGTGCCAAAGTGGA 2106
 959 ... 969
 2107 GCGCATTTGATTAAGCAATCACGCC...CAAGCATTTT 2144
 970 glyaspleuasnlethrnsnglulysasprthrgluetelele 986
 2145 CGGTGTCGACCCATCAAGCCACACAACTGTACAGCTTCGACTGA 2194
 986 yglyasprvalserglulysglulysnleuthrleiserasplyst 1003
 2195 CGGCTGTGACAGTTGTACCAAAAACATTTC...GACGATAAAGTG 2241
 1003 leasnlethr...lyslethrlethrlelysalaglyal 1015
 2242 ATTGCTTCATTAAGCAACGACGACATGAGGCAATGTCCGTCGGA 2291
 1016 aspylgluasnseraspserraspalaThrAsnAsnalaasnleuthr 1032
 2292 TCACGCT...CATTAATCTCACAGACTT... 2319
 1032 elythrlysluasnleuthrlnaspleuasnleiserdlyphea 1049
 2320 ...GCCACACTCAACGCG 2334
 1049 snlysalaglylethrAlalyasprlyseraspleuthrlelyasn 1065
 2335 AATCTTAGTGAGGCGGAGACGACACTATAGCGGTACGCGCAAGCGCAC 2384
 1066 thrasnserAlaasprglythrAsnAlalyssvalthrphasnlnva 1082
 2385 CCAAAACGCGCAACCTCAGCTCGTGGCAATGCGCAACATTTATC 2434
 1082 llyaspserylleserAlaasprglyhlysalvalthrleuhsler 1099
 2435 AAGCCACATTAAAGCGCAACACATCGCTGGCAATGCTCATTTAT 2484
 1099 ysvalgluthrserglyserasnasnlnthrlnaspserraspasn 1115
 2485 CTAACCAACAGCGCGTACAAACGCGAGCTGAGCTTTCGACACGCG 2534
 1116 asnalaglyleuthrleaspralalyasnvalthrvalaasnasn 1132
 2535 TAAGCAAC...GTAAGCATTTCCGACATCAAGCGCATGTC... 2574
 1132 ethrerhlysalavalserlleserAlaThrserglyleulthr 1149
 2575 ...TCCGTAGCGGTAAGCGAGTTCATTTGAAACAGC 2613
 1149 hrllysthrglythrthrleasnalaThrthrglyasnvalglulthr 1165
 2614 CGCTTACCGGAATAATCAGCGGCGCAAGATACG...GCAT 2654
 1166 AlaglnthrlyserleuclglyleugluserSerSerlyser 1182

APPLICATION NUMBER: US/08/728,470

512 CAGAACCTGTTGAG.....ATGACCAGTTATATGGAT 543

303 **ta**lysgluGluAlaGluIleGlyGlyValIleSerAlaGlnAsnGln 319

544 GGTGGAATACGCTGATTAAATAATACCTGATCTGT.....CG 587
 520 GlnAlaIysGlyLysLeuMetIleThrGlyAspIysValThrLeu 336
 588 AATCGGACA.....GGCAGCAATATT 610
 336 sthGlyAlaValIleAspLeuSerGlyLysGlyLysGlyLysThr 353
 611 GCGGCTCATGAA...GACGACCAATAAC..... 639
 333 euGlyGlyAspIuArgGlyLysGlyLysAsnGlyIleGlnLeuAlaLys 369
 640CGGAAAGTTCATATCATATTCACAC..... 666
 370 LysThrSerLeuGlyLysGlySerThrIleAsnValSerGlyLysGly 386
 667GCATATCTTG.....CTGCTGGTGGCAATA 694
 386 sGlyIyArgAlaIleValITrpGlyAspIleAlaLeuIleAspIysAsnI 403
 695 CTTTGCACAAAATGATCAGGTGGTGACAGTCACACTAGTAGCGAA 744
 403 IeAsnAlaGln...GlySerGly..... 409
 745 AAAATTAACATAGCCCATATGCTTTTACCACAGAGGCTCATTTGG 794
 410AspIleAlaLysThrGlyLysPheValG 419
 795 GCACAGTGGCTCACCAATGTTATCTATGATGCCCAAGCAAAAGTGT 844
 419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
 845 TAATTAATGGGTATTCACACAGCAACCCCTATATAGAAAAGCAAT 894
 430AsnAlaIleValAspAla..... 435
 895 GGCCTTCACACTAGTTCGTAAAGATTGTTCTATGAAATCTTTGCTGG 944
 436LysGlnTrpLeuAsp.....Ph 442
 945 AGATACCCATTCAGTATCTACGACACATCAAAATGGAAATACCTTT 994
 442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
 995 TTACACACATTAATAAGCGCAGCAAAATGATGCCAAACATAAACAC 1044
 459 IuAspAspGluTrpThrGlySerGly.....AsnSer 469
 1045 TATTCCTACCTATAGATTAAACACAGAACCTTCATTTGTTAATGT 1094
 470 AlaSerThrProLysArgAsnLysGlyLysThr...LeuThrAsnThr 485
 1095 TTCTTATTCGAG.....ACAGCAA 1114
 485 rThrIleuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
 1115 GAGAACTGTTAT..... 1128
 502 snGlnAlaGlyIyValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
 1129CATGCTGACGTGGGTCAACAGTTATCGACC 1160
 519 ThrLeuTrpSerGluGlyArgSerGlyGlyVal..... 530
 1161 CAGACTGAATAT.....GAGAAATA 1183
 531 GlnIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsn 547
 1184 TTTCCTTATTCACAAAGAAAGTGAATGATTAATTACACAGACATC 1233
 547 euThrIleTrpSerGlyLysTrpValAsp.....ValHisLysAsnIle 561

1234 AACCAAGCGCGCGCGGCTTTTATTTTGAAGGCTAAATTTACGCTCCGC 1283
 562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
 1284 TAAA.....AACACGAAGAGTGGCAAGCGC 1309
 573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleTrpGlyG 590
 1310 CCGGCTTCATATACGTGATGGCAGTACCGTTACTTGGAAAGTAAACGCG 1359
 590 IuGlyThr...IleTrpSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
 1360 GTGGCAAAAGCACCGCCTGTCCAAATGCGCAAGGACCGCTGCTGTTCA 1409
 606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTr 618
 1410 AGCCAAAGGGAAC.....CAAGCTCGG 1435
 618 rThrLysArgThrAsnLysTrpAlaIleThrAsnLysPheGluGlyThr 635
 1436 TCAGCGTGGCGGACGTAAGTCATCTTA.....GAT 1467
 635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
 1468 CAGCAGCGGACGATCAAGCAAAACAAACCTTT..... 1503
 651 GluSerGlyTrpAspLysPheLysGlyArgThrTrpAsnLeuThrSe 667
 1504AGTGAATCGGC.....TTGCTGACGCGCAGG 1531
 667 rLeuAsnValSerGluSerGlyLysPheAsnLeuThrIleAspSerArg 684
 1532 GGAG.....GTG 1539
 684 LysSerAspSerAlaGlyThrLeuThrGlnProTrpAsnLeuAsnGlyIle 700
 1540 CACTGAATCCCGATATCACTTCAACCCGACAAA..... 1575
 701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
 1576CTCT 1579
 717 nPheAspIleLysAlaProIleGlyIleAsnLysTrpSerSerLeuAsn 734
 1580 ATTTCGGCTTTCGCGCGGACGTTTGATTTGAACGGGCAATCGCTTCG 1629
 734 yAlaSerPheAsnGlnLysAsnIleSerValSerGlyGlyLysValAsp 750
 1630 TTCACCGCATT.....CAAAATCCGATGAAGGCGGATGATGT 1670
 751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValAla 767
 1671 CAACCAAT.....CAAGCAANAG 1690
 767 eAsnSerLysTrpPheAsnValSerThrGlySerSerLeuArgPheLys 784
 1691 AATCCACGTTATCCATTACAGC.....AATAAGATATTAATCT... 1728
 784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
 1729 ...ACACCGGCATATAC..... 1743
 801 AsnAlaThrGlyLysAsnIleThrLeuLeuGlnValGluGlyThrAspG 817
 1744AACACTTGATAGCAAAAGAAATGCTCTCAACGCT... 1782
 817 yMetIleGlyLysGlyIleValAlaLysAsnIleThrPheGluGly 834
 1783TGCTTGGCGGAAGATGCA...ACCAAAACGAACGCGCG 1821
 834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
 1822 CTCATATGTAATTACCAACCGCAAGAGCGGATCGCACTTA..... 1863

```

881 ValThrIleAsn.....AsnAsnAlaAsnValThrIleuIleGlySer 864
1863 ..... 1863
864 rAspHeaSpasnHsIsglnLysProIeuThrIleLysLysAspValIle 881
1864 .....CTGCTTCCGGCGCA.....ACAATTAAACGGC 1893
881 leaSerSerGlyAsnLeuThrAlaGlyLysnIleValAsnIleAlaGly 897
1894 AATTATCNG...CAACAAACGGCAAACTGTTTTCAGCGCA..... 1933
898 AsnLeuThrValGluSerAsn..AlaAsn...PheLysAlaIleThrAsp 913
1934 .....GACGACACCGCAGCGCTTACAAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysSglLysnSerAsnIle 929
1958 ATTAGAGACGGGCTGTCAAAATGAGAGTATCCACAGAGAGAAATC 2007
930 SerIleAlaLysGlyLysAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGCGGACACAGATTGATGATGACCGC..ACATTAAACGGCAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerThrTyArg..... 958
2057 ATATTACGGGCGGACACGCGTGTCCCGCAATGTTGCCAAAGTGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GGGGATTGGCATTTAAGCAATCACGCC.....CAAGCATTTT 2144
970 GlyAlaPheAsnLleThrAsnGlnLysSerAspThrGlnMetGlnIleG 986
2145 CGGTGTGCGACCGCATCAAGCACACCAATCTGTACAGCTTGGAGCTGA 2194
986 yGlyAspValSerGlnLysGlnLysLeuThrIleSerSerAspLysI 2003
2195 CGGGTCTGACAGTGTACCGGAAAAAACCATTC...GAGCATTAAGT 2241
1003 leaSnIleThr.....LysGlnIleThrIleLysAlaGlyAla 1015
2242 ATGTCTTCATTGACAGACGACGACATCAAGGCAATGTCAGCTTGCCGA 2291
1016 AspGlyGlnAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrI 1032
2292 TCACGCT.....CAATTAAATCTCACAGAGCTT. 2319
1032 eLysThrLysGlnLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049
2320 .....GCCACACATCAACGGC 2334
1049 snLysAlaGlnIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTCAGCGCGAGACACGACATATACGGTTACGGCGCACGGCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnI 1082
2385 CCAAAACGGCAACTCAGCTGTGGCGAATGCCAAGCAACATTTATC 2434
1082 lLysAspSerLysIleSerAlaAspGlyHsLysValThrLeuHsSerI 1099
2435 AAGCCACATTAAAGCGCAACACATCGGCTTCGCAATGCTTCATTTAAT 2484
1099 ySValGlnThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAGCAACAACGCCGTACAAAGGACGATCGAGCTTCCGACACAGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysValThrValAsnAsnAsnI 1132
2535 TAAAGCAAC.....GTAAGCATTCGCGACGTCACAGCGCAATGTC.... 2574

```

```

1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThr 1149
2575 .....TCCTAGCCGATTAAGCAGTATTCATTTGAAAACAGC 2613
1149 hLysThrGlyThrThrIleAsnAlaThrThrGlnValGluIleThr 1165
2614 CGCTTACCGGAAAAATCAGCGCGCGCAAGATACG.....GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyIleGluSerSerSerGlySerI 1182
2655 ACACCTTAAAGACAGCGAATGACGCTGCCGTGGGACAGGAATTAAGCA 2704
1182 lThrLeuThrAlaThrGlnLysAlaLeu.....AlaValSerA 1195
2705 ATTTAAACCTTGAACAGCGCACCATTAACATTCATTCGCTATGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGCGCAGCGCGCAACCGGACGTGCGGACAGATGCGCGCGCGCG 2804
1208 .....GlyAlaLeuThrThrIleuAlaGlySer..... 1216
2805 TTGCGCGCTTCCCTTATTCGCTTACGCGCGCACTTTCGAGAAATCC 2854
1217 .ThrIleLysGlyThrGluSerValThr.....ThrSerSerGlnSer 1231
2855 GTTTCACACACGCTGACGTTAAACGGCAATTGAACGTCACAGACA... 2901
1231 Ly.....AspIleGlyThrIleSerGly...GlyThrVal 1242
2902 ....TTCCGCTTATGCGGAGACTCTGCGCTACCGCAGCGCAATTTAA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnLysIleL 1259
2949 GCTGCGGAAAGTCCGAGCAGCTTACACCTTGCGCTTC..... 2988
1259 s...AlaThrThrGlnGlyAlaAsnValThrSerAlaThrGlyThrIleG 1275
2989 .....AACATACCGGCAACGAACCCGTAAGTCTGAGCA 3024
1275 LyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGly 1291
3025 TTGACGGTA.....GNGAGAGAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGlnIleAsnAlaThrGlnGlyAlaIa 1308
3051 CACACCGCTGTCCGAAATCTTAATTCACCGTCAAAAGCAACACGTCG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGlnAlaSerSerHisIle 1325
3101 ATCGCGCGCATGGCGTTATGAGCTTATCCGCAAAAGACGGCAGTCCGC 3150
1325 hSer..AlaLysGlyGlnValAsnLeuSerIleAlaIleAspIleSer 1341
3151 CTGCATTAATCCGGTCAAGAAAGCAAGAGCTTCCGACA 3187
1341 aGlySerIleAsnAlaIleAsnValThrLeuAsnThr 1353

```

seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:US-08-617-697-2

seq_documentation_block:

```

; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlaire, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; SUITE: Bldg. 1
; CITY: Arlington

```



```

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

```

```

alignment_scores:
Quality: 265.00      Length: 1297
Ratio: 0.438         Gaps: 73
Percent Similarity: 46.646      Percent Identity: 21.126

```

alignment_block:
US-09-303-518D-653 x US-08-617-697-2 ..

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

```

178 AATAAGGCAAGTTTGCACGCGGCGAAGATATGAGCTTACACAA 227
    ||| ||| :|||:||||| ||||| :|||
199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211
228 AAAAGGGAGTTGTGCGCAATCGATGACGAAGCCCGATGATTGATT 277
    :|||:||||| :|||: ||| :||| :|||
212 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 221
278 TTCTGTGTATCGCGTAAAGGC.....GTGCGGCATTGGCGGCGCAT 321
    :|||:||||| :|||: ||| :||| :|||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237
322 CAA.....:|||||:|||||:|||||:|||||:|||||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGGCAGATTAAGCGCGCTATTAACAATGTTTGTGCGGAGGAA 385
    :|||:||||| :|||: ||| :||| :|||
254 AlaIleProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAAT.....:|||||:|||||:|||||:|||||:|||||
271 LysAsnIleAsnValAlaGluAlaIleThrIleArgAsnGlnGlyLysLeu 287
412 TACCAATTGTGAAGAAGAATAATTATTAAGCAGGACATAACGCCATCC 461
    :|||:||||| :|||: ||| :||| :|||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297

```

```

462 TTATGGCGGCAATTATCATATGCCGCTTGGCAAAATTTGTCAAGATG 511
    :|||:||||| :|||: ||| :||| :|||
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTGAG.....ATGACCGATTATATGAT 543
    :|||:||||| :|||: ||| :||| :|||
303 IalysGluGlyGluAlaGluIleGlyGlyValIleSerValIleAsnGln 319
544 GGGTGAATATAGCCGATTAAATAATACCCGATCGTGT.....CG 587
    :|||:||||| :|||: ||| :||| :|||
320 GluAlaIleGlyGlyLysLeuIleThrIleThrGlyAspLysValThrLeu 336
588 AATCGAGCA.....GCCAGACAAATATT 610
    :|||:||||| :|||: ||| :||| :|||
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGluThrTyrL 353
611 GCGGCTGTATGAA...GACGAACCAATATAC..... 639
    :|||:||||| :|||: ||| :||| :|||
353 euGlyGlyAspGluArgGlyGlyLysAsnGlyIleGluLeuAlaLys 369
640 .....CGCGAAAGTTCATATCATATTCGAACG..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlu 386
667 .....GCATATTTCTTG.....CTCGTGGTGGCAAT 694
    :|||:||||| :|||: ||| :||| :|||
386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaIleAspGlyAsnI 403
695 CCTTGGCAAAATGCAATCAGTGGTGGCGACAGTCAACTAGTACGAA 744
    :|||:||||| :|||: ||| :||| :|||
403 LeuAlaIleGln...GlySerGly..... 409
745 AAATTAACATAGCCCATATGTTTATTCACACAGAGGCTCATTGG 794
    :|||:||||| :|||: ||| :||| :|||
410 .....AspIleAlaLysThrGlyGlyPheValGlu 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCAAACCAAAAGGT 844
    :|||:||||| :|||: ||| :||| :|||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGATTTGCAACAGCAACCCCTATATAGGAAAGCAAT 894
    :|||:||||| :|||: ||| :||| :|||
430 .....AsnAlaIleValAspAla..... 435
895 GCCTTCACAGTAGTTCGTAAGATTTGTTCTATGATCAATCTTCTGG 944
    :|||:||||| :|||: ||| :||| :|||
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGATTTCTACGAACCCATCAAAATGGAAATACTTT 994
    :|||:||||| :|||: ||| :||| :|||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSer 459
995 TTAACGACAAATAATATGCGCGAGGAAATAATCGATGCCAAACATAAC 1044
    :|||:||||| :|||: ||| :||| :|||
459 LuAspAspGluTrpThrGlySerGly.....AsnSer 469
1045 TATTCCTACCTTATATGATTAATAACAGCAACCGTCAATGTTTAATGT 1094
    :|||:||||| :|||: ||| :||| :|||
470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnThr 485
1095 TTCTTATATCCGAG.....ACAGCA 1114
    :|||:||||| :|||: ||| :||| :|||
485 ThrIleuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTTAT..... 1128
502 sngLArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129 .....CATGCTGAGGTGGGTCACACAGTTATCGACC 1160
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVal..... 530
1161 CAGACTGATTAAT.....GGAGAAATA 1183

```

```

531 .....|||
    GluLeuAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTGACAAAGGAAAGTGAATTGATCTTACCAAGCAATC 1233
    .....|||
547 eThrIleThrSerGlyThrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGCTTGTATTGAGGTATTTTACGGTCCGCC 1283
    .....|||
562 SerLeuGlyAla.....GlnLysAsnIleAsnIleThrAl 573
1284 TAA.....AACCAAGCAAGCTGGCAAGCG 1309
    .....|||
573 aLysGlnAspIleAlaPheGlyLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGGGTTTCATATCATGATGACAGTACCGTACTTGAAAGTAAGCAAGC 1359
    .....|||
590 IngIlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAAGCAAGCGCGCTGTCCAAATCGCAAGCAAGCGCTGGTGTCA 1409
    .....|||
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheThr 618
1410 AGCCAAAGGGAAGAAC.....CAAGCTCGG 1435
    .....|||
618 rThrLysArgThrAsnLysThrAlaIleThrAsnLysPheGlnGlyThrL 635
1436 TCACGGTGGGCGGCGTAAAGTATCTTA.....GAT 1467
    .....|||
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGGAGCATCAAGGCAAAAAACAAGCCTT..... 1503
651 GluSerGlyThrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....AGTGAATCGGC.....TTGTCACGCGCAGGG 1531
667 rLeuAsnValSerGluSerGlyLupheAsnLeuThrIleAspSerArgG 684
1532 GGACG.....GTG 1539
684 LysSerAspSerAlaGlyThrLeuThrGlnProLysAsnLeuAsnGlyIle 700
1540 CAACGTGAATGCCGATATCATGATCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGGTTTCGCGCGAGCGATTGATTTGAAGGGCATTCGCTTCG 1629
    .....|||
734 yAlaSerPheAsnGlnLysAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGCATTT.....CAAAATACGATGAAGGGCGCATGATGT 1670
    .....|||
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValValI 767
1671 CAACCAACAT.....CAAGACAAG 1690
    .....|||
767 eAsnSerLysThrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATTCACCGTTTACCATTTACAGC.....AATAAAGATATTACT... 1728
    .....|||
784 hrSerGlySerThrLysThrGlyPheSerIleGlyLysAspLeuThrLeu 800
1729 ...ACAACCGCAATAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspG 817
1744 .....AACAACCTTGATACCAAAAAAGAAATTGCTTACACAGCT. 1782
    .....|||
817 yMetIleGlyLysGlyIleValAlaLysAsnIleThrPheGlnGlyG 834
1783 .....TGCTTGGCGAGAAAGATGCA...ACCAAAAGCAAGCGCGG 1821
    .....|||
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGlnGlyAsn 850
1822 CTCATCTGAATTATCAACCGGAGAACCGGATCGCACTTTA..... 1863
    .....|||
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysAspValIleI 881
1864 .....CTGCTTCCGCGCGA...ACAAATTAAACGCG 1893
    .....|||
881 LeuAsnSerGlyAsnLeuThrAlaGlyLysAsnIleValAsnIleAlaGly 897
1894 AATATCACG...CAACAAAGCGCAAACTGTTTTCAGCGGCA..... 1933
    .....|||
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934 .....GACCGACACCGCACGCTTACATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTAGGAACCGGCTGTGTCAAAATGGAAGGTATCCCAAGGAGAAATC 2007
    .....|||
930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGACACACGATTTGATTCAGCCGCACATTTAAAGCGGAAACTTCC 2056
    .....|||
946 nLeuSerIleThrThrAsnSerSerThrTyrArg..... 958
2057 ATATTCAAGGCGGACAAAGCGGTGTTCCCGCAATGTTCCAAAGTGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGGATTTGGCATTTAAGCAATCAGCGC.....CAAGCAGTTT 2144
    .....|||
970 GlyAspLeuAsnIleThrAsnGlnGlySerAspThrGluMetGlnIleG 986
2145 CGGTTCGACCGCATCAAGCCACACAAATGTCACAGTTGGAGTCGA 2194
    .....|||
986 yGlyAspValSerGlnLysGlnGlyAsnLeuThrIleSerSerAspLys 1003
2195 CGGCTGTACAGATTGTACCGGAAAAAACATTAC...GACGATAAAGTG 2241
    .....|||
1003 LeuAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGAGCAAGCAGACATCAGAGCAATGTCACGCTTGGCGA 2291
    .....|||
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrI 1032
2292 TCACGCT.....CATTTAAATCTCACAGCACTT. 2319
    .....|||
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049
2320 .....GCCACAGTCACAGCG 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTGAGCGGAGACACGACGACATACGTTACGCGCAAGCGCAC 2384
    .....|||
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGln 1082
2385 CCAAAAGGCAACGTCACGCTGTGGCAATGCCCAAGCAACATTTAATC 2434
    .....|||
1082 LysAspSerLysIleSerAlaAspGlnLysValThrLeuHisSerI 1099
2435 AAGCCACATTTAAAGCAACACATCGCTTGGGCAATGTTATATTAT 2484
    .....|||
1099 yValGluThrSerGlySerAsnAsnThrGluAspSerSerAspAsn 1115

```

```

2485 CTAGCAACAACGCGGTACAAAAGCGACGCTGACGCTTTCGCAACAGC 2534
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1116 AsnAlaGlyLeuThrIleAspAlaValThrValAsnAsnAla 1132
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2535 TAAGGCAAC.....GTAGCCATTCCGACCCACGCAATGTC.... 2574
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThr 1149
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2575 .....TCCCTAGCCGATTAAGCAGATATTCATTTTGAACAGC 2613
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1149 hTrpThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluIleThr 1165
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2614 CGCTTACCGGAAATATCAGCGCGGCGAGATACG.....CGATT 2654
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGlnSerSerGlySerVal 1182
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2655 ACACCTTAAGACAGACGATGACGCTGCGGCGAGATGAGCA 2704
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1182 IThrLeuThrAlaThrThrGluGlyAlaLeu.....AlaValSerA 1195
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2705 ATTTAAACCTTGACAAAGCCACCATTCATTCCTGCTATGACAC 2754
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2755 GATGGCGGAGCGGCGCAACCGCGGCGAGATGCGCGCGCGCGCG 2804
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2805 TTCGCGCGCTTTCCTTATTCGCTTACCGCGCACTTCGCGAGATCC 2854
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1217 .ThrIleGlyThrGlnSerValThr.....ThrSerSerGlnSer 1231
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2855 GTTTCAACACGCTGACGCGTAACGCGCAATTAAGGTCAGGACAA... 2901
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2902 ...TTCGCTTATGTCGAGACTCTCGGCTACCGCGGCGCAATTCGA 2948
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1243 GluValLysAlaThrGlnSerLeuThrThrGlnSerAsnSerLysIle 1259
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2949 GCTGCGGGAAGTTCGGAAGGACATTCACCTTGCTGCTC..... 2988
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1259 s...AlaThrThrGlyLysValAsnValThrSerAlaThrGlyThrIle 1275
      : : : : : : : : : : : : : : : : : : : : : : : : : :
2989 .....AACATACCGGCAAGAACCCGTAGTCTCGAGCA 3024
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1275 LysIleThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGly 1291
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3025 TTGACGCTA.....GTGGAAGGAAGAACAA 3050
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaThr 1308
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3051 CACACCGCTGCGCAAAATCTTAATTCACCTGCAAAAGCAACAGCTGC 3100
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIle 1325
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3101 ATGCGCGGCGATGCGCTTATCAGCTTATTCGCAAGGCGGAGTTCGCG 3150
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1325 hrSerAlaLysGlyGlnValAsnLeuSerAlaGlnspcIlySerVal 1341
      : : : : : : : : : : : : : : : : : : : : : : : : : :
3151 CTGCAATATCCGCTCAAGAACAGACCTTCGAGCA 3187
      : : : : : : : : : : : : : : : : : : : : : : : : : :
1341 agLysrIleAsnAlaAsnValThrLeuAsnThr 1353
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:US-08-719-641-2

```

seq_documentation_block:
: Sequence 2, Application us/08719641
: Patent No 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins

```

```

: TITLE OF INVENTION: OF NO. 6218141-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Spomaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berksstresser, Jerry M
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: US-08-719-641-2

alignment_scores:
Quality: 265.00 Length: 1297
Ratio: 0.438 Gaps: 73
Percent Similarity: 46.646 Percent Identity: 21.126

alignment_block:
US-09-303-518D-653 x US-08-719-641-2 ..
Align seg 1/1 to: US-08-719-641-2 from: 1 to: 1536

178 AATAAGCAAGTTCGAGTCGCGGAAAGATATTGATTGCAACAA 227
      ||| ||| : : : : : ||| ||| ||| : : : |||
199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211
      : : : : : : : : : : : : : : : : : : : : : : : : : :
228 AAAAGGAGATTGTCGCAATTCGATGACGAAGCCCGATGATTGATT 277
      : : : : : : : : : : : : : : : : : : : : : : : : : :
212 .....AsnLeuIleGlyGlyLysValLysAsnLys..... 221
      : : : : : : : : : : : : : : : : : : : : : : : : : :
278 TTTCTGTGTATTCGCTTAACGCG.....GTGCGGCAATTGGCGGCGAT 321
      : : : : : : : : : : : : : : : : : : : : : : : : : :
222 ..GlyValIleSerValAsnGlyLysrIleSerLeuLeuAlaGlyGln 237
      : : : : : : : : : : : : : : : : : : : : : : : : : :
322 CAA.....TATATTGTGAG 335
      : : : : : : : : : : : : : : : : : : : : : : : : : :
238 LysIleThrIleSerAspIleLeuAsnProThrIleThrTySerIleAl 254
      : : : : : : : : : : : : : : : : : : : : : : : : : :
336 CGTGCACATAACGCGGCTATACCAATGTTGTTGTCGCGAGGAGAA 385
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAAT .....CCGATCAGACCGCTTCT 411
271 LysAlaIleAsnValArgAlaIleThrIleAlaGlnGlnGlyLysLeuSer 287
412 TACCAATGTGAAGAATAATTATTAAGCAGGACTAACGGCCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGCGATTATCATATGCGCGTTGCACCAATTTGTCAAGATG 511
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTTGAG.....ATGACCACTTATATGAT 543
303 LAlysGlnGlyGluAlaGlnIleLeuGlyAlaIleSerAlaGlnAsnGln 319
544 GGGTGAATAATACGCTGATTAAATTAATACCTGATCGTGT.....CG 587
320 GlnAlaLysGlyLysLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGAGCA.....GGCAGCAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGlnGlyGlnThrTyrL 353
611 GCGCGCTGTGATGA...GACGACCCATATAC..... 639
353 euGlnGlyAspGluArgGlyGlnGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGCGAAGTTCATATCATATGTCAGC..... 666
370 LysThrSerLeuGlnLysGlySerThrIleAsnValSerGlyLysGlnLys 386
667 .....GCATATTCTTGG.....CTCGTGGTGCATA 694
386 sGlyArgAlaIleValTyrGlyAspIleAlaLeuIleAspGlyAsnI 403
695 CTTTGCACAAATGATCAGTCAGGTGGGAGCAGTCACCTAGTAGCCGA 744
403 LeAsnIleGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGTTTTCACACAGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTATGGGTATTCACACAGCAGCAACCCCTATATAGAAAAAGCAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCCTTCACAGTACGTTGCTTAAGATGGTTCTATGATGAATCTTGTGG 944
436 .....LysGlnTyrPheLeuAsp.....Phe 442
945 AGATACCATTCATGATCTACGACACATCAAAATGGGAATACCTTT 994
442 eAspAsnValSerIleAsnAlaGlnThrAlaGlyArgSerAsnThrSerG 459
995 TTACAGACAAATTAATGCGCAGAGAAATATCATGCCAACAAATCAACAC 1044
459 LuAspAspGlnTyrThrGlySerGly.....AsnSer 469
1045 TATTCCTACCTTATAGATTAAAAACAGCAACGCTCATTTGTTAATGT 1094
470 AlaSerThrProLysArgAsnLysGlnLysThrThr...LeuThrAsnTh 485
1095 TTCTTTATCGAG.....ACAGCAA 1114
485 rThrLeuGlnSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502

1115 GAGAACCTGTTTAT..... 1128
502 sGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129 .....CATGTCAGAGTGGGGTCAACAGCTATATGACC 1160
519 ThrLeuTyrSerGlnGlyArgSerGlyGlyVal..... 530
1161 CAGACTGAAATAT.....GGAGAAATA 1183
531 GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTCACAAAGAAAGGATGATGATTCTTACCAACATC 1233
547 euThrIleTyrSerGlyGlyTyrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGCGGCTTGTATTGAGGATTAATTACGGTCGCC 1283
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAA.....AACACGAAACGTGCAAGCG 1309
573 aLysGlnAspIleAlaPheGlnLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGCGCTTCATACAGTACGAGTACGCTTACTTGGAAATACAGGC 1359
590 InGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAs 605
1360 GTGGCAACAGCACCGCCTGTCCAAATCGCAACAGCACGCTCTGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheThr 618
1410 AGCCAAAGGGGAAAC.....CAAGCTCGG 1435
618 rThrLysAlaGlyThrAsnLysTyrAlaIleThrAsnLysPheGlnGlyThrL 635
1436 TCAGCGTGGGCGACGGTAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CACGAGCGGAGATCAACGCAAAACACACCTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTyrPheAsnLeuThrSe 667
1504 .....AGTGAATCGGC.....TTGTCACGCGGACGG 1531
667 rLeuAsnValSerGlnSerGlyGlnPheAsnLeuThrIleAspSerArg 684
1532 GGACG.....CTG 1539
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACGATGGCGGATATACATTCATCAACCCGACAAA..... 1575
701 SerPheAsnLysAspPThrPheAsnValGlnArgAsnAlaArgAlaLys 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGCTTTTCGCGCGGACGTTGGATTGAAACGGGCAATTCGCTTTCG 1629
734 yAlaSerPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGCATT.....CAAAATACCGATGAAGGCGCATGTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValAlaI 767
1671 CAACCACAAT.....CAAGACAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784

```



```

1341 aGlySerIleAsnAlaAsnValThrIleuAsnThr 1353
seq_name: /cgn2_6/prodata/1/1aa/6B_COMB.pap:US-09-268-347-36
seq_documentation_block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

alignment_scores:
Quality: 257.50      Length: 1701
Ratio: 0.341         Gaps: 75
Percent Similarity: 44.444      Percent Identity: 18.871

alignment_block:
US-09-303-518d-653 x US-09-268-347-36
Align seg 1/1 to: US-09-268-347-36 from: 1 to: 2411

7 ACAACGCAAAAGCGACACCGAA.....ACACCGCAAGCCCC 47
||| |||||:||||: ||| :|||
242 ThrGlyAspLysAsnThrIleuAsnValIleuThrAlaLysGluAsnG 258
||||| :|||:||||: ||| :|||
48 TAAACGGCGCCATCCGCTTCGCCCGCTTACTTACCAATATGCGCCGT 97
||||| :|||:||||: ||| :|||
258 LysThrThrGluValLysPheThrProLysThrSerValIleuLysGlu 275
||||| :|||:||||: ||| :|||
98 CGTTGGCGATTCTGCCCAAGCGCGGAGACACACTTATTTGGCGATC 147
||| ||| ||| :|||: ||| :|||
275 YAspGlyLysLeuPheThrGlyLysGluAsnAspThrAsnLysVal 291
||||| :|||:||||: ||| :|||
148 AACTACCAATCTATCGGACTTGGCCCAATAATAAGCAAGTTGCAGT 197
||| :||| :|||: ||| :|||
292 ThrSerAsnThrAlaThrAspAsnThrAspGluLysAsnGlyLeuVal 308
||||| :|||:||||: ||| :|||
198 CGGGCGCAAAAGATATTGAGTTTACACAAAAGAGGAGTGGTCGCA 247
||||| :|||:||||: ||| :|||
308 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTyrPArgValLys 325
||||| :|||:||||: ||| :|||
248 AATCATACGCAAAAGCCCGCATGATGATTT...TCTGTGTATGCGGT 294
||| :||| :|||: ||| :|||
325 hrThrAlaAsnGlyLysAsnGlyLysPheAlaThrAlaLysSerGly 341
||||| :|||:||||: ||| :|||
295 AACGGCGTGGCGCATTTGGCGGATCAATATATGTGAGCGCGGACCA 344
||||| :|||:||||: ||| :|||
342 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLys 358
||||| :|||:||||: ||| :|||
345 TAACGGCGCTATAACAATGTTGATTTGGTGGAGGAGCAATCCCG 394
||| :||| :|||: ||| :|||
358 sAspThrAsnGlyAsnGly..... 364
||||| :|||:||||: ||| :|||
395 ATCAGCAACCGCTTTTACCAATTTGTGAAGAAATAATTAAGA 444
||| :||| :|||: ||| :|||
365 ..... 373
||||| :|||:||||: ||| :|||
445 GGAAGTACGGCCATCTTATGCGGGGATATCATATGCGGCTTGA 494
||| :||| :|||: ||| :|||
374 Gly...AspGlyLeuLysPheAspSerAsp.....Lys 383
||||| :|||:||||: ||| :|||
495 CAAATTTGTCAAGATGCAAGACCTGTGATGACAGTATATGATG 544
||||| :|||:||||: ||| :|||

```

```
635 rSerLySergLu.....AsnGlyLySHist 644
1274 CGGTCGCGCTAAACAAACGACGAGCGCGCGCTTCATATC 1323
      |||
644 hTlEthrValSerValAlaGluThrLysAlaAspSerGlyLeuGlu... 659
      |||
1324 AGTATGGCAGTACCGTTACTTGGAAAGTAAGCGCTGGCAACGACCG 1373
      |||
660 LysAspGlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAs 676
      |||
1374 CCTGTCCAAATCGGC...AAGGCGAGCTGCTGTTCAGGCCAAAGGG 1420
      |||
676 nValLeuThrValGlyAsnAsnGlyThrAlaValThrLysGlyLysPheG 693
      |||
1421 AA.....AACCAAGCTCGCTGAGCGTGGCGAGCGTAAGTC...ATC 1461
      |||
693 luthrValLysThrGlyAlaThrAspAlaAspArgGlyLysValThrVal 709
      |||
1462 TTAGATCAGACGGCGGACGATCAAGGCAAAACACAGCC...TTTAGTGA 1508
      |||
710 LysAspAlaThrAlaIleAsnAspAlaAspLysLysValAlaThrValLys 726
      |||
1509 AATCGCGCTTGTCACGGCGGAGGGGAGCGTGCACACTGAATGCCGATAT 1557
      |||
726 pValAlaThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnL 743
      |||
1558 .....CAGTTCAACCCC.....GAC 1572
      |||
743 euThrThrSerLysAspGluAspAsnProThrAspAsnGlyLysAspAs 759
      |||
1573 AAACCTCTATTTCGC.....TTTGGCGGCGGA.....CG 1601
      |||
760 AlaLeuLysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLys 776
      |||
1602 TTTGGATTGAACGGGCGATTCGCTTTCGTTCCACCGCATTCCAAAATACCG 1651
      |||
776 sValLysAlaGlyAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuG 793
      |||
1652 ATGCAAGGGCGGATGATTGTCAACCAACATCAAGACAGATCCACCGTT 1701
      |||
793 lValLysThrAlaLysValSer.....AspThrLeu 803
      |||
1702 ACCATTACGCGCAATAA..... 1719
      |||
804 ThrIleGlyLysAsnThrProThrGlyGlyThrAlaThrProLysVal 820
      |||
1720 GATATTACTACACCGGCAATTAACAACACTTG..... 1752
      |||
820 lAsnIleThrSerThrAlaAspGlyLeuAsnPheAlaLysGluThrAla 837
      |||
1753 .....GATACCAAAAAAGAAATTGCCTACACCGT.....TGTTT 1788
      |||
837 sPalaSerGlySerLysAsnValTyrLeuLysGlyIleAlaThrLeu 853
      |||
1789 GCGGAGAAAGATGCAACCAACCAAGACGGCGCTCAATCTGAATTACCA 1838
      |||
854 ThrGluProSerAlaGlyAlaLysSerSerHisValAspLeuAsnValAs 870
      |||
1839 ACCGGAAGAACGGATCGCACTTAA.....CTGCTTCCGGCG 1876
      |||
870 pAlaThrLysLysSerAsnAlaIleSerIleGluAspValLeuArgAlaG 887
      |||
1877 GAACAATTTAAACGGCAAT..... 1896
      |||
887 lYrPaSnIleGlnGlyAsnGlyAsnAsnValAspTyrValAlaThrTyr 903
      |||
1897 .....AT 1898
      |||
904 AspThrValAsnPheThrAspSerThrGlyThrThrValThrIva 920
      |||
1899 CACGCAACA..... 1908
      |||
920 lThrGlnLysAlaAspGlyLysGlyAlaAspValLysIleGlyAlaLysT 937
      |||

1909 .....AACGCAAACTGTTTTCAGCGCAGACCG 1938
      |||
937 hSerValIleLysAspHisAsnGlyLysLeu...PheThrGlyLysAsp 952
      |||
1939 ACACCGCAGCGCTACATCATTTAGAGACGGGTGCTCAAAATGAGACG 1988
      |||
953 LeuLysAspAlaAsnAsn.....GlyAlaThrValSerGluAspAsp 967
      |||
1989 TATCCACAGAGAGAA...ATGCTGGGACAAACGTTTGATCGACCGCA 2035
      |||
967 yLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleAspAlaV 984
      |||
2036 CATTTAAAGCGGAAACTTCATATTCAGGCGGACACGCGTGTTC 2085
      |||
984 alAsnLysSer...GlyTPArgValThrGlyLysValAlaThrAlaGlu 999
      |||
2086 CGCAATGTTGCCAAAGTGAAGGCGATTGGCATTTAAGCAATCACGCCA 2135
      |||
1000 ThrGlyAlaThrAlaVal.....AsnAlaGlyAsnAlaGluTh 1012
      |||
2136 AGCAGTTTTCGCTGCGCA.....CCGCAATCAAGCACACACATCT 2176
      |||
1012 rValThrSerGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThra 1029
      |||
2177 GTACACGTTGCGACTGACGCGGT..... 2199
      |||
1029 lathrValSerLysAspAsnGlyAsnIleAsnValLysTyrAspValAsn 1045
      |||
2200 .....CTGACAGCTTG 2210
      |||
1046 ValGlyAspGlyLeuLysIleGlyAspAspLysLysIleValAlaAspTh 1062
      |||
2211 TACGCAAAAAACCATTCACCGCATTAAGTATGATTCCTTACGACAGA 2260
      |||
1062 rThrThrLeuThrValThrGlyLysValSerValProAlaGlyAla 1079
      |||
2261 CCGACATCAGAGGCATGTCAGCCTGCCATCAGCTCATTTAAATCTC 2310
      |||
1079 snSerValAsnAsnLysLysLeu.....ValAsnAla 1090
      |||
2311 ACAGACTTGCACACATCAACGCGCAATCTTACT..... 2343
      |||
1091 GluGlyLeuAlaThrAlaLeuAsnAsnLeuSerThrPheAlaLysAlaAs 1107
      |||
2344 .....GCAGGCGGAGACGCGCACTTAACGCTT... 2370
      |||
1107 pLysTyrAlaAspGlyLysSerGluGlyLysThrAspGlnGluValLysA 1124
      |||
2371 .....ACGCGCAACGCCACCCAAACGCAACCTCAGCTGTGGCGCAT 2415
      |||
1124 lAglyAspLysValThrPheLysAlaGlyLysAsnLeuLysValLysGln 1140
      |||
2416 GCCCAAGCAACATTTAAT.....CAAGCAGATTAACGCGCAACG 2456
      |||
1141 SerGluLysAspPheThrTyrSerLeuGlnAspThrLeuThrGlyLeuTh 1157
      |||
2457 ATCGGCTTCGCAATGCTTCATTTAATTAAGCAACAC.....G 2497
      |||
1157 rSerIleThrLeuGlyGlyThrAlaAsnGlyArgAsnAspThrGlyThv 1174
      |||
2498 CCGTACAAAACGCGAGTCTGACGCTTCGCAACAGCTAAGGCAACGTA 2547
      |||
1174 alIleAsnLysAspGlyLeuThrIleThrLeuAlaAsnGlyAlaAla 1190
      |||
2548 AGCCATTCGCACTCAACGCAATGCTCCCTAGCGGATTAAGCAGATAT 2597
      |||
1191 GlyThrAspAlaSerAsnGlyAsnThr..... 1199
      |||
2598 CCAATTTGAACACGCCGCTTAACCGGAAAAATCAACGCGCGGCAAGATA 2647
      |||
1200 .....lSerValThrLysAsp. 1205
      |||
```

```

2648 CGGCATTACCTTAAAGACAGCAATGACGCTCCGCGGACAGGAA 2697
1206 .....GlyIleSer 1208
2698 TTAGCAATTTAAACCTTGACAAACGACACATTACACTCAATTCGCCCTA 2747
1209 AlaGlySerLysGluThrAspValLysSerAlaLeuLysThr...Ty 1224
2748 TCGA.....CAGCATGCGGACGCGGCAAAACGCGAGTGGCGAGGATG 2791
1224 rLysAspThrGlnAsnThrAlaGlyAlaThrGlnProAlaAlaAsnThr 1241
2792 CGCGCGCGCGCGCTCGCGCGCTTCCTATTATTCGCTACGCGCGCACT 2841
1241 La.....GluValAlaLysGlnAspLeuValAspLeuThrLysProAla 1255
2842 TCG.....GCAGATCCCGCTTCAACACGCT 2867
1256 ThrGlyAlaAlaGlyAsnGlyAlaAspAlaLysAlaProAspThrThrAl 1272
2868 GACGCTAAACGCGCAATTTGACGCGTCAAGCAATTCGCTTTATGTCGG 2917
1272 aaLathValGlyAspLeuArgGlyLeuGly..... 1282
2918 AACTCTTCGCTACCGGACGCAATTTGACGCTGCGGAAAGTTCGCA 2967
1283 .....TrpValLeuSerAlaLysLysThrAlaAspGluThrGlnAsp 1296
2968 GGCATTTACACCTTGGCTGTCAACATTTACCGGCAAC...GAACCGCTAG 3014
1297 LysGluPheHisAlaAlaValLysAsnAlaAsnGluValGluPheValGlu 1313
3015 TCCTGAGCAATTTGACGCTAGTGGAAAGAAAGCAACACACCGCTGTCG 3064
1313 yLysAsnGlyAlaThrValSerAlaLysThrAspAsn..... 1325
3065 AAATATTTATTTCAACCTGCAAAAGCAACGCTGATGCCGCGGCATGG 3114
1325 ..... 1325
3115 CGTTATACCTTATCCGCAAGACGCGAGTTCGCTCATTAATCCGCT 3164
1326 .....AsnGlyLysHisThrValThrIleAspVal 1335
3165 CAAGACAAAGACTTCCGCAAACTGCGCAAGCGGAGCAAGCAAGC 3214
1335 LAlaGluAlaLysValGlyAspGlyLeuGluLys.....AspThrAspG 1350
3215 CGCCTTGACGCGCAAA..... 3231
1350 LysValIleLysLeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThr 1366
3232 ...CAGGCAACAATTGCGCGCAAAACAGCGGAAAAAGCAACGCGCA 3278
1367 ValAspAlaThrLysGlyAlaSerValAlaLysGlyLupheAsnAlaVala 1383
3279 AAGCCTTACGCGCTGATTCGCGCGGCGCAATGCCCGCAAAAGCAG 3328
1383 LThrThrAspAlaThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyL 1400
3329 AAAGTGTTCGCAACGCGCGCGGACGCGAGGCAAAATGCCGCGCAT 3378
1400 yValValValLys.....GlySerAsn...GlyAla 1409
3379 ATGACGCGGCGGAGAAAGAAAGAGGCTCAGCGGAGTAAAGACAGCG 3428
1410 ThrAlaThrGluThrAspLysLysValAlaLThrValGlyAspValAla 1426
3429 CTTGGCGCAACAGCGCGCAAGCG.....GAACCGCGCGCG 3463
1426 alYsAlaIleAsnAspAlaAlaThrPheValLysValGluAsnAspAspS 1443
3464 CTACACACGCGCTTCGCCGCGCGCGCGCGCGCGGATTTGCCGCA 3513
1443 ..... 1451
3514 CGCAGAGCCCAACCGCAACCCGCAACGCGGACGCGCTATACCGCTTA 3563
1452 AspAspGlyAlaAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLys 1468
3564 TGCCATATAGCGGTTTGAAGTAATTTCCGCCAGCGCTCAACAGCGTTTCG 3613
1468 sAlaGlyLysAsnLeuLysValLysArgAspGlyLysAsnIleThrPheA 1485
3614 CCGTACAGAGCAATTTGACCGCGGATTCGCGGAGACCGCGGCAAGCC 3663
1485 lAlaLeuAlaAsnAspLeu.....Ser 1491
3664 GTTGGACAGCGCATCCGCGGACACCAACACTACCGTTGCGAGATT 3713
1492 ValLysSerAlaThrValSerAsp..... 1499
3714 CGCGGCTTACCGCAACACCGACCTGCGCAATTCGATGCAAGAA 3763
1500 .....LysLeuSerLeuGlyThrAsnGlyAsnValAla 1511
3764 ACCTCGCAGCGGCGCGCTGCGCATCTGTTTCG...CACACCGGACC 3810
1511 snLleThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThr 1527
3811 GCA.....ACACCTTCGACGAGCGCATCGGCAACTCGGACGCT 3851
1528 GlyAspAspAlaAsnIleHisLeuAsnGlyLleAlaSerThrLeuThr... 1543
3852 TGGCCAGGTCGCTTTCGGGCAATACGGCATGCGAGTTCACATCG 3901
1544 .....AspThrL 1546
3902 GCATACGCGCGGCGCGGTTTGTAGTACGCGACGCTTTCAGACGATC 3951
1546 euLeuAsnSerGlyAlaThrThrAsnLeuGly.....GlyAsnGlyLle 1560
3952 AAGCAAAATTCGCGCGCGCGCTGCTGATTCGCGCATTCAGCGCAAGTA 4001
1561 ThrAspAsnGluLysArgAlaAla.....SerValLysAspValLe 1575
4002 CGCGCAGGTTTCGCGGATTCGCGCATCGCAACCGCATCGCGCAAGC 4051
1575 uAsnAlaGlyThrAsnValArgGlyValLysPro..... 1586
4052 GCTATTTGTCCAAAAGCGGATTACGATACGAAACGTCATATTCGCC 4101
1587 .....AlaSerAlaAsnAsnGluValGluAsnIleAspPheVal 1599
4102 ACCCGGCGCTTGCAATTCACCGCTACCGCGCGGCAATTAAAGCAGATTA 4151
1600 Ala..... 1600
4152 TTCATTAAACCGCGGCAACACATTTCATCAGCGCTTATTGAGCCTGT 4201
1601 .....ThrTyra 1603
4202 CCTTACCGATGCGCTTCGCGCAAAAGTCCGAAGCGGCTCATATACGCC 4251
1603 sPThrValAspPheValSerGlyAspLysAspThr...ThrSerValThr 1618
4252 GTATTGCGGAGATTTTCGCGCAAAACCGGACGCGGATGGCGGCTTAA 4301
1619 ValGluSerLysAspAsnGlyLysArgThrGluValLysIleGlyAlaLys 1635
4302 CGCGGAATC.....AAAGTTCACAGCTGTCCCTTC 4333
1635 sThrSerValIleLysAspHisAsnGlyLysLeuPheThr..... 1648
4334 ACCTGCGCGCGCAAGGCGCAATTTGGAAGCGGACGCGGCGC 4383

```



```

611 Asn.....GlyThrLysGluLeuSerAsn...GlnVa 620
1278 CTCGCTTAATAAACAGCAAGGCTGCGCAAGCGCGCGCTTCATATCAGTG 1327
      ::::: ::::: |||||
620 LlysGlnAlaIleAspGlnValLeuPheThrGlyAlaGly..... 632
1328 ATGGCAGTACCGTTACTTGGAAAGTAAAGCGCTGCGCAACGACCGCTG 1377
      ::::: |||||
633 ..AlaAlaThrValThrSerLys..... 639
1378 TCATAAATCGCAGCAAGCGCTGCTGTTCAAGCCAAAGGGAACCA 1427
      ||||| |||||
640 SerGlnAsnGlyLysHisThrIle..... 647
1428 AGGCTCGGTAGCGTGGCGCAGCGTAAGTCATCTTAGATCAGACGGCG 1477
      ::::: |||||
648 ...ThrValSerValAlaGlnThrLys..... 655
1478 ACGATCAAGGCAAAAAAAGCCTTAGAGTAATCGGCTTGTGACGGCG 1527
      ::::: |||||
656 .....AlaAspGlyLeuGlnLysAsp 663
1528 AGGGGAGCGTGCACACTGAATGCCGATTAATCAAGTTCACCGCAAACT 1577
      ||||| |||||
664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680
1578 CTATTTCGGCTTCGCGCGGACGCTTGGATTGAACGGGACATTCGCTT 1627
      ||||| :::::
680 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGlnThrV 697
1628 CGTTCACCGCATTCAAATACCGATGAGGGCGCATGATCTGTACAC... 1674
697 alLysThrGlyAlaThrAspAlaAspArgLysValThrValLysAsp 713
1675 .....CACATCAAGACAAGAA... 1692
714 AlaThrAlaAsnAspAlaAspLysValAlaThrValLysAspValAl 730
1693 .....TCCACCGTTACCATTAACGCAATTAAGATTAATTA 1729
730 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGlnAsnLeuThrT 747
1730 CAAC...GGCAATACACAACACTGATAGCAAAAAGAATATGCGTAC 1776
      ||||| :::::
747 hSerIleAspGlnAspAsnProThrAspAsnGlyLysAspAlaLeu 763
1777 AACGGTTGG.....TTTGGCGAAGAAAGATGCAACAA 1808
      ::::: |||||
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLy 780
1809 AACGAACGGGCGG.....CTCATCTGCAATTAAC 1837
      ||||| :::::
780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGlnVal 797
1838 AACCGAAGAAAGCGATGCACTTACTGCTTCCGCGCAACAATTA 1887
      ::::: |||||
797 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 813
1888 AACGCG.....AATATCGCAAAACAACGCA 1916
      ::::: |||||
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspG 830
1917 ACTGTTTTCAGCGCAGACCGACCGACCGCTTCAATCAATCAATTTGGA 1966
      ||||| :::::
830 yLeuAsnPheAla...LysGlnThrAlaAspAla.....SerGlyS 843
1967 GCGGGTGTCAAAAATGGAAGGTATCCCAAGAGAAATCGTGTGGAC 2016
      ||||| :::::
843 eLysAsnValLysLeuLysGlyLeuAla..... 852
2017 AACGATTGATGACCGCACATTTAAAGCGGAAACCTCATATTAG... 2064
      ::::: |||||
853 ThrThrLeuThrGlnProSerAlaGlyAlaLysSerSerHisValAspLe 869

```

```

2065 .GGCGGACAAGCGGTGTTCCCGCAATGTTCGCAAGTGA... 2106
      ::::: |||||
869 uAsnValAspAlaThrLysSerSerAsnAlaAlaSerIleGlnAspVal 886
2107 .....GGCGATTTGGCATTTAAGCAATCACGCCCAACACATTTTCGTC 2151
      ||||| :::::
886 euAlaGlyAlaGlyTTPAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 902
2152 GCACCGCATCAAGCCACACATCTGTACAGCTTCGACGTGACGGGTCT 2201
      ||||| |||||
903 AlaThrLysAsp.....ThrValAsnPheThrAspSerThrGlyTh 917
2202 GACAAGCTTACCGAAAAAACCTTACGACGATAA.....GTGATG 2245
      ||||| |||||
917 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 934
2246 CTTCATTTGACGACGACGACATC..... 2268
934 ySrlleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950
2269 .....AGAGCAATGTACAGCTTGGCGA 2291
951 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrThrValSerGlnAs 967
2292 TCACGCTCATTTAATTCACAGAGACTTCCGCACTCAACGCAATCTTA 2341
      ||||| :::::
967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 984
2342 GTGCAGCGCGAGACACGCACTATACGTTTACG...CGCAACGCCACCA 2388
      ||||| :::::
984 sPlalValAsnLysSerGlyTTPArgValThrGlyGlnGlyAlaThrAla 1000
2389 AACGGAACCTCAGCCTCTG.....GGCAATGCCAAGCAACATTTAA 2432
      ::::: |||||
1001 GlnThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGlnThrValThrSe 1017
2433 TCACGCACTTA.....AACGGCAAC.....ACATCGG 2461
      ||||| :::::
1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1034
2462 CTTCGCAATGCTTCATTTAATCTAAGCAACAGCGCTAACAACGGC 2511
      ::::: |||||
1034 eLysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyAsp 1050
2512 AGCTGAGCGTTTCGACACAGCTAAGCAACGTAAGCAATTCGCGACT 2561
      ::::: |||||
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrIle 1067
2562 C.....AACGCAATGTCTCCCTA..... 2580
      ||||| :::::
1067 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1084
2581 .....GCCGATAAGCGAGTATTCATTTTGA 2607
      ||||| :::::
1084 sAsnAsnLysLysLeuValAsnAlaGlnGlyLeuAlaThrAlaLeuAsn 1100
2608 AACAGCGCTTTACCGGAAAA..... 2628
      ||||| :::::
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyGlnSerI 1117
2629 .....ATCAGCGCGCGC..... 2640
      ||||| :::::
1117 uGlyGlnThrAspGlnGlnValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AAGATACGCGCATTAAC 2658
      ||||| :::::
1134 lAglyLysAsnLeuLysValLysGlnSerGlnLysAspPheThrLysSer 1150
2659 TTAAAGACAGCAATGAGCGTGCCTGCG.....GGCAGCA 2696
      ||||| :::::
1151 leuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167

```

```

2697 ATTAGCAAT.....TTAACTTGACAGCCACA 2728
      |||
1167 aasnlglyargasnphrghlythrvalileasnlyaspglyleuthrli 1184
      |||
2729 TTACACTCAATTCGGCTATCGACAGATCGCGGAGCGGCAACCGGC 2778
      |||
1184 lethrleu.....alaasnlglyalaalaagly 1193
      |||
2779 AGTCGGAGATGCGCGCGCGCTTGGCGCGTTCCTATATTCGT 2828
      |||
1194 thraspha.....serasnlyasnthrliaseriva 1204
      |||
2829 TACGCGCCCAACTTCGGCAGATCCGTTTCACACGCTGACGGTAAACG 2878
      |||
1204 lthr.....lysaspglyleaserlaaglyasnlysgluilethrnsn 1219
      |||
2879 GCAATTCAGCGGTACGAGAACATTCGCTTATGTCGAGACCTTCGGC 2928
      |||
1219 alyls.....seralaleulysthr 1225
      |||
2929 TACCGCAGCGCAATTCAGCTGCGGAAAGTTCCGAAAGCACTTACAC 2978
      |||
1226 tyrllyasphthrghlnasnthrvalasapgluthrghlnasplysgluphe 1242
      |||
2979 CTGGCTGTCAACAATTCACCGGCAAC..GAACCGTAACTCTCGACGCAAT 3025
      |||
1242 salalaivalyllyasnalaasnlglyalgluphevalgllylyasnlgly 1259
      |||
3026 TGACGGTGTGAGAGAAAGACACACACCGCTCCGAAAACTTAAAT 3075
      |||
1259 latthrvalserlalysalysphasn..... 1267
      |||
3076 TTCACCTTCGAAACGACACGTCGATGCGCGCATGCGCTTATCACT 3125
      |||
1267 ..... 1267
      |||
3126 TATCCGCAAGAGCGGAGTTCGCGCTGATTAATCCGTCAAAGAACAG 3175
      |||
1268 .....asnlglyshisthrvalthrleaspyalalagluhial 1281
      |||
3176 AGCTTCCGACAAACTTCGCAAGCGGGAAGAAAGAGCGCGCTTGAGC 3225
      |||
1281 ysvalgllyaspglyleuglylys.....AspThrAspIlystlelys 1295
      |||
3226 GCAAAA.....CAGGCACA 3239
      |||
1296 leulysvalaspasnthraspglyasnasnleuThrValaspalath 1312
      |||
3240 ACTTGGCGCAACACAGCGGGAAGAAAGACGCGCAAGCGCTTGAGC 3289
      |||
1312 rlysglyalaservalalalysgllyupheasnalaivalthrThAspA 1329
      |||
3290 CGCTGATTCGCGCGCGCAATCCACGAAAGAGAGAAACTGTTCG 3339
      |||
1329 lathrthrleaglnlythrAsnalaasnlglylyvalalval 1345
      |||
3340 GAACCGCGCGGACGAGCGGGAATGCGCGCATATGACGCGGCA 3389
      |||
1346 lys.....glyserasn..glyalathralathrgl 1355
      |||
3390 GGAAGAAAAAAGGGGTGAGCGGATAAAGACACCGCTTGCGCAAC 3439
      |||
1355 urthraspIlylyalvalalathrvalgllyaspvalalalysalale 1372
      |||
3440 AGCGCGAGCG.....GAACCGCGCGCTTACACCGCC 3474
      |||
1372 snasphalaalathrPheVallyvalglusnaspsaspheralathrile 1388
      |||
3475 TTCCCCCGCGCGCGCGCGCGCGGATTTGGCGCAACGCGCGCCA 3524
      |||
1389 .....AspAspSerProthrAspAspIly 1397
      |||
3525 ACCGAAACCCCAACGCGAGCGCATGATCAGCGCTTATGCCAATAGCG 3574

```

```

1397 aasnspalaleulysalaglyAspThrleuthrleulysalaglyysA 1414
      |||
3575 GTTTGAGTAATTTTCGCCACCTCAACACGCTTTGGCGCTACAGAC 3624
      |||
1414 snleulysvallylsargaspIlylyasnliethrPheleuAlaasn 1430
      |||
3625 GAATTGACCGCGTGTTCGGAAGACCGCGCAACGCGCTTGGAACAG 3674
      |||
1431 Aspleu.....serVallylsSerAl 1437
      |||
3675 CGGCATCCGGAGACCAACACTACCGTTCGCAAGATTTCCGCGCTCAC 3724
      |||
1437 athrvalaserasp..... 1441
      |||
3725 GCCAACAACCGACCTCGCCCAATCGTATGACGAAACCTCGGACG 3774
      |||
1442 .....lysleuSerleuglythrAsnlglyasnlyvalasnliethrser 1456
      |||
3775 GGGCGGTCGCGCATCCTGTTTCG...CACAACCGAGCGGA..... 3813
      |||
1457 AspThrlysglyleuysPhealalysaspserlysthrGlyaspal 1473
      |||
3814 .AACACCTTCGACGAGCGCATCGCAACTCGGACGCGCTTCCCAACGTG 3862
      |||
1473 aasnliethleuasnlglylealaserthrleuthr..... 1485
      |||
3863 CCGTTTCGGCAATACGCGATCGGCGATGTCGACATCGGCATCAGCGC 3912
      |||
1486 .....AspThrleuLeuAsnser 1491
      |||
3913 GCGCGGCTTTTAACTAGCGGACGCGCTTTCAGACGCGATCGACCAAT 3962
      |||
1492 glyalathrthrleuasnleugly.....glyasnlglylethrAspasnlg 1506
      |||
3963 CCGCGCGCGCTGCTGCTATACGCGCATTCAGGCAAGATACCGGAGGT 4012
      |||
1506 ulyslystrglaala.....serVallyaspvalleuasnliagly 1521
      |||
4013 TCGGCGATTCGCGCATCAACCGCACATCGCGCACGCGCTATTTGTC 4062
      |||
1521 rpsasnvalargglyalyspro..... 1528
      |||
4063 CAAAAGCGGATTCGATACGAAAGCTCAATATCCACCCCGGCGCT 4112
      |||
1529 AlaserAlaasnlnvalgluasnliethrPhevalala..... 1542
      |||
4113 TGCAATTCACCGCTACCGCGCGCATTAAGCAGATATTCATCAAC 4162
      |||
1542 ..... 1542
      |||
4163 CGGCGCAACATTTCCATCACGCTTATTTAGCGCTGCTATACGAT 4212
      |||
1543 .....ThrTyAspThrValasp 1548
      |||
4213 GCGCTTCGCGCAAGTCCGAAGCGCGCTCATATACCGCGGTATGGCGCA 4262
      |||
1549 PhevalserglyAspIlyAspThr...ThrserValthrValgluSerly 1564
      |||
4263 GGAATTTCGCAAAACCGCAGTGGCGAATGGGCGCTTAACCGCAATC 4311
      |||
1564 saspsasnlglyshrglythrghluvallystlelyalalysthrservali 1581
      |||
4312 .....AAAGTTTCACGCTGCTCCCTCACGCTGCGCC 4344
      |||
1581 lelyasphlasnlglyleuPheThr..... 1590
      |||
4345 GCCAAGGCGCGCAATTCGAACGCAACGACGCGGCGGATC 4386
      |||
1591 .....GlylysgluLeuLysaspAlaasnasnlglyval 1602
      |||

```

seq_name: /cgn2_6/ptodata/1/1aa/6B_COMB_pdp:US-09-377-155-33

[illegible]

```
1278 CTCGCCCTAAAAACACAGACGTCGCAAGCGCGCGCTTCATATCATGTG 1327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
620 llysglnalaspigluvalleuphetrhlgllyalgly..... 632
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAAAGACCGCTG 1377
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
633 ..AlaIaIaThrValThrSerLys..... 639
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1378 TCCAAATATGGCAAAAGCAGCTGCTGTTCAAGCCAAAGGGAAACCA 1427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
640 SerGluasnnglyLysHisThrIle..... 647
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1428 AGGCTCGGTACGCTGGCGGACGGTAAAGTCATCTAGATCAGAGGGG 1477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
648 ...ThrValSerValAlaGluThrLys..... 655
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1478 ACGATCAAGCAAAAAACACAGCTTGTAGTAATCGGCTGTGTCAGCGC 1527
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
656 ..... 663
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1528 AGGGGACGCTGCACATGATCGCATATCATGTTCAACCCCGCAAACT 1577
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
664 GlyAspThrIleLysLeuLysValaspasnnglnsnThrAspAsnValle 680
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1578 CTATTTCGCTTTTCGGCGGACGCTTGTGATTGAAAGCGCATTCGCTTT 1627
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
680 uthrValGlyasnnglyThrAlaValThrLysGlyGlyPheGluThrV 697
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1628 CTTTCACCGCATTCAAATACCGATGAAAGGCGCATGATTGTCAAC... 1674
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
697 alLysThrGlyAlaThrAspAlaAspArgLysValThrValLysasp 713
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1675 .....CACATCAAGACAAAGAA..... 1692
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
714 AlaThrAlaAsnaspAlaAspLysValAlaThrValLysaspValAl 730
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1693 .....TCCACGCTTACCATAGCAATTAAGATTATCTACTA 1729
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
730 aThrAlaIleasnSerAlaIaIaThrPheValLysThrGluAsnLeuThr 747
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1730 CAAC...GGCAATAACAACACTTGGATAGCAAAAGAAATTCCTTAC 1776
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
747 hSerIleAspGluAspAsnProThrAspAsnnglyLysAspAspAlaLeu 763
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1777 AACGGTTGG.....TTGGCGGAGAAAGATGCAACCA 1808
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVal 780
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1809 AACGACGGGGCG.....CTCATCTGAATTACC 1837
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluVal 797
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1838 AACCGGAAGACGATCGCACTTACTCTTCCGCGGAAACAATTTA 1887
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
797 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 813
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1888 AACGGC.....AATTCACGCAACAACAACGGCA 1916
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAsp 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1917 ACTGTTTTCAGCGGACGACACCGACGCGCTTCAATTCATTATAGAA 1966
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
830 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 843
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1967 GCGGGGTGCTAAAAAATGAAGTATCCCAAGAGCAAAATCGTGTGGAC 2016
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
843 erLysAsnValTyLeuLysGlyLea..... 852
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2017 AACGATTGGATCGACGCGACATTAAGCGGAAACCTTCATATTCAG... 2064
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
853 ThrThrLeuThrGluProSerAlaGlyAlaLysSerHisValAspLe 869
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2065 .GGCGGCAAGCGGTGGTTCCGCAATGTTCGCAAAAGTGAA..... 2106
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
869 uasValaspAlaThrLysLysSerAsnAlaAlaSerIleGluAspVal 886
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2107 ....GGCGATTGGCATTTAAGCAATCAGCCCAACGACCTTTGGGTGC 2151
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
886 euArgAlaGlyTrpAsnIleGlnnglyAsnnglyAsnAsnValAspTyVal 902
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2152 GCACCCATCAAAAGCCACACAACTCTGTACAGCTTGGAGTGGAGGCTCT 2201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
903 AlaThrThrAsp....ThrValAsnPheThrAspAspSerThrGlyTh 917
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2202 GACAACTTGTACGCAAAAAACATTACGACATTA.....GTGATG 2245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
917 rThrThrValThrValThrGluLysAlaAspGlyLysGlyAlaAspVal 934
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2246 CTTCATTGACGACGCGACATC..... 2268
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
934 yIleGlyAlaLysThrSerValIleLysAspHisAsnnglyLysLeuPhe 950
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2269 .....AGAGCAATGTGAGCCTTCCGA 2291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
951 ThrGlyLysAspLeuLysAspAlaAsnAsnnglyAlaThrValSerGlu 967
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2292 TCACGCTCATTTAAATCTACAGACTTGCACACTCAGCGCAATCTA 2341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 984
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2342 GTGAGCGCGGAGACACGCACTATACGCTTACG...CGCAAGCGCACCA 2388
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
984 spAlaValAsnLysSerGlyTyTrpArgValThrGlyGluGlyAlaThrAla 1000
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2389 AACGCAACCTCAGCCTCTGTC.....GGCAATGCCCAACCAATTTAA 2432
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1001 GluThrGlyAlaThrAlaValAlaAsnAlaGlyAlaAsnAlaGluThrVal 1017
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2433 TCAAGCCACATTA.....AAGGCAAC.....ACATCGG 2461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1017 rGlyThrSerValAsnPheLysAsnnglyAsnAlaThrThrAlaThrVal 1034
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2462 CTTGGCAATAGCTTCATTAACTAAGCAAAACGCGTACAAACGGC 2511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1034 erLysAspAsnnglyAsnIleAsnValIlystyryAspValAsnValGly 1050
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2512 AGTCTGACGCTTTCGACAAACGCTAAGCAACGTAAGCCATTCGCGACT 2561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 GlyLeuLysIleGlyAspAspLysIleValAlaAspThrThrThrIle 1067
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2562 C.....AAGGCAATGTCTCCCTA..... 2580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1067 uThrValThrGlyLysValSerValProAlaGlyAlaAsnSerVal 1084
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2581 .....GCCGTAAGGACGATTCATTCATTGAA 2607
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1084 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2608 AACACGCGCTTACCGGAAA..... 2628
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyryAlaAspGlyLysSer 1117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2629 .....ATCAGCGGCGG..... 2640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLys 1134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2641 .....AAGCATACGCGATTACAC 2658
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1134 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTySer 1150
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2659 TTTAAAGACAGCGAATGAGCGCTGCGCTG.....GGCAGCA 2696
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2697 ATTAGGCAT.....TTAACTTGACAAAGCCACCA 2728
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```

1167 aasnlglyargasnspthrnglythrvalleasnlylaspolythrthri 1184
2729 TTACACTCAATTCGCCCTATCGACACGATGGCGGCGGCAACCGCG 2778
1184 lethrleu.....Alaasnlglyalaalaalagly 1193
2779 AGTGGCGGAGATGCGCGCGCGCGCGCTTCGCCCTTATTCATTCGT 2828
1194 thrspala.....Serasnlglyasnthrleleerva 1204
2829 TACGCGCGCACTTCGGCAGATCCCTTTCACACGCTACCGGTAAAG 2878
1204 lthr.....LysaspglyileseralaglyasnlylsgluilethrAsnV 1219
2879 GCAATTTGAACGCTCAGGGAACATTCGCTTATGTGCGAATCTTCGCG 2928
1219 allys.....Seraleuylsthr 1225
2929 TACGCGCGGCAATTTGAAGCTGGCGGAAGTTCCGAAAGCACTTAC 2978
1226 TylrlysnspthrGlnasnthrAlaaspGluThrGlnasplysgluheh 1242
2979 CTGGCTGTCAACAATACCGCAAC...GAACCGTAACTTCGAGCAT 3025
1242 salalavallyllyasnalaasnlglyalgluphevalglylysasnGlyA 1259
3026 TGACGCTAGTGGAGGAAAGAACACACACCGCTGTCCGAAATCTTAAT 3075
1259 lathrvalSerAlaValThrAspAsn..... 1267
3076 TTCACCTGCAAAACGAAACACGTCGATGCCGCGCATGCGTTATCAGCT 3125
1267 ..... 1267
3126 TATCCGAAAGCGCGGATTCGCTGCATATCCGCTCAAGAACAG 3175
1268 .....Asnlglyshisthrvalthrleasvalalagluvalal 1281
3176 AGCTTCGCAACAATCGCGCAGCGGCGGGAAGAACAGAGCCCTTGACG 3225
1281 ysvalglyaspolyleuGluLys.....AspThrAspPolylysileys 1295
3226 GCAAAA.....CAGGCACA 3239
1296 leulysvalaspsnthraspGlyasnleuleuthrvalaspAlaTh 1312
3240 ACTTGCGCGCAACACAGCGGAAAGAACACCGCAACCTTGACG 3289
1312 rlysglyalaservalalalysglupheasnalaValThrThrAspa 1329
3290 CGCTGATTCGCGCGCGCATGCGACGCAAGAAAGCAGAAAGTGTGCC 3339
1329 lethrthrAlaGlnGlythrAsnAlaasnlglylvalValVal 1345
3340 GAACGCGCGCGCGCAGCGGGAATAATGCGCATTTATGACGCGGA 3389
1346 lys.....glyserasn...GlyAlaThrAlaThrGln 1355
3390 GGAAGGAAAGAAACGCTGCGAGCGGATTAAGACACCGCTTGCGCAAC 3439
1355 uthrAspLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1372
3440 AGCGCGGAAGCG.....GAACCGCGCGCGCTACACACCGCC 3474
1372 snAspAlaAlaThrphValLysValGluAsnAspAspSerAlaThrIle 1388
3475 TTCGCCCGCGCGCGCGCGCGCGGATTTGCGCAACCGCAGCCCA 3524
1389 .....AspAspSerProthrAspAspGlyAl 1397
3525 ACCGCAACCCCAACGCGCGCGCATGATGACCGTTATGCGCAATGCG 3574
1397 aAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGlyLysA 1414
3575 GTTTGAGTGAATTTCCGCCACCGCTCAACGCGTTTTCGCGGTACAGAC 3624
1414 snleuLysVallylArgAspGlyLysAsnThrlethrphleAlaAsn 1430
3625 GAATTGACCGCGCTGTTTGGCGGAAGACCGCGCAACCGCTTGGCAAG 3674
1431 Aspleu.....SerVallylSerAl 1437
3675 CGGCATCCGGGACACCAACACTACGTTTCGCAAGATTCGCGGCTAC 3724
1437 aThrValSerAsp..... 1441
3725 GCCAACAACCGACCTGCGCAATCGGTATGCAAGAAACCTGCGACG 3774
1442 .....LysleuSerleuGlyThrAsnGlyAsnlylValasnThrSer 1456
3775 GCGCGCGTCCGATCCGTTTCG...CAACCGGACCGGA..... 3813
1457 AspThrLysGlyleuAsnThrAlaLysAspSerLysThrGlyAspAspAl 1473
3814 AACACCTTCGACGACGCGATCGGCACTGGCGACGCTTGCCACGCTG 3862
1473 aAsnIleHisLeuAsnGlyIleAlaSerThrleuThr..... 1485
3863 CCGTTTCCGCGCAATACGGCATCGGCGATTCGACATCGGCTCAGCGG 3912
1486 .....AspThrleuLeuAsnSer 1491
3913 GCGCGCGGTTTATGATGCGGACGCTTCAGACGCGCATCAGCAAAAT 3962
1492 GlyAlaThrThrAsnleuGly.....GlyasnGlyIleThrAspAsnG 1506
3963 CCGCGCGCGCGTCTGCTGATTCAGGCTTACGCAAGATACCGCGAGTT 4012
1506 uLysLysAlaGlyAla.....SerVallylAspValleuAsnAlaGlyT 1521
4013 TCGGCGATTCGCGCATCGAACCGCACATCGCGCAACGCTATTTGCTC 4062
1521 rpsnValArGlyValLysPro..... 1528
4063 CAAGAACGCGATTCGCGATGCAAGAACGTCATATGCGACCGCGGCT 4112
1529 AlaSerAlaAsnAsnGlnValGluAsnIleAspPheValAla..... 1542
4113 TGCATTCACCGCTACCGCGCGGCTATTAAGCAGATTATTCATTCAAC 4162
1542 ..... 1542
4163 CGGCGCAACATTTTCATCAAGCGCTTATTTGAGCTGTCTATACCAT 4212
1543 .....ThrTyrAspThrValAsp 1548
4213 GCCGCTTCGCGCAAGTCCGAACGCGCTCAATACCGCGATTGGCGCA 4262
1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerly 1564
4263 GGATTTCGCAAAACCCGACAGTGGCGAATGGGCGTAAAGCGCAATC. 4311
1564 sAspAsnGlyLysArgThrGluValLysIleGlyAlaLysThrSerVal 1581
4312 .....AAAGTTTCAAGCTGTCTCCATCAAGCTGCGCGC 4344
1581 leLysAspHisAsnGlyLysLeuPheThr..... 1590
4345 GCGAAGGCGCGCAATTTGGAAGCGCAGCAGACCGCGGCTAC 4386
1591 .....GlyLysGluLeuLysAspAlaAsnAsnAlaVal 1602

```

seq_name: /cgn2_6/prodata/1/laa/6B_COMB.pep:us-08-913-942-4
seq_documentation_block:
; Sequence 4, Application US/08913942

```

; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; City: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1-RFT/RMS/DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELE: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-913-942-4

alignment_scores:
    Quality: 253.50      Length: 1664
    Ratio: 0.344         Gaps: 75
    Percent Similarity: 44.231    Percent Identity: 19.591

alignment_block:
    US-09-303-518d-653 x US-08-913-942-4 ..

Align seg 1/1 to: US-08-913-942-4 from: 1 to: 2353

7 ACAACCGACAAGACGACACGAA.....ACACACCGCAAGAGCCC 47
   ||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:
246 Thrlglyspgysasnhrleuaspvalvalleuthrlalalsglsnsl 262
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 TAAACCGCGCATCCGCTTTCGCGCCGCTTACTAGCATATGCGCTG 97
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 ylsthrthrlvallysphehrprolysthrservalilleysglut 279
98 CGTTCGGCATTCGCCCCAAGCCGCGGCGGACACACTATTTTCGGATC 147
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 ysaspglylyslleuphethtlylsqslusnhsnaspthralslyval 295
148 AACTACCAATATCGGACCTTTCGCAAAATAAAGCAAGTTGCGACT 197
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 Thrsersnhralathraspsnhraspplucllyasnsllyleuvalth 312
198 CGGGCGCAAGATATGAGCTTTCACAAACAAAGAGGAGTTGGTCGCA 247
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

312 ralatysalavalilleasplavalasnllysalaglyttrpargvallst 329
248 AATCGATGACGAAGCCCGATGATGATTT...TCGTGGTATGCGCT 294
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 hrthrhralaaasnglyglnasnglyasphealahrvalalasergly 345
295 AACGGCGTGGCGCATTTGGCGGCGATGATATATTTGAGCGTGGACA 344
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
346 ThrAsnValthrhegluserglyaspglythrThrAlaserValthrly 362
345 TAAAGCGCGCTATACAAATGTTGATTTGGTCGGGAGGAGCAATCCG 394
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 saspthrasnsllyasnslly..... 368
395 ATGACGACCGCTTTCTTACCAATTTGTGAAGAATAATATTAAGCA 444
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 .....llethrvallystyralsalysval 377
445 GGGACTACGCGCATTCCTTATGCGGCGATTCATATGCGCGTTTGA 494
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 Gly...Aspglyleuylsphenaspsersp.....ly 387
495 CAATTTGTACAGATGACAGACCTGTTGAGATGACCGATTATATGATG 544
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 slslyllevalalaspthrthrAlaleuthrvalthr.....G 400
545 GGTGAATACGCTGATTAATTAATACCTGATCGTGTGCAATC... 591
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
400 lyglylyvalalagluillealalsglnaspsaplyslslyleuval 416
592 .....GGAGCAGCGACGACATATTTGGCG... 615
417 AsnalaglyaspleuvalthrAlaleuglyAsnleusertrpysalaly 433
616 .....TCGTGATGAGAC.....GAACCAATACCGCAAGATT 649
433 salagluAlaspthrhraspglyalaleugllylleserlysaplng 450
650 CATATCATATTTGCAAGCGCATTTCTTGGCTGCTGCGTGCAAT...ACC 696
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
450 luvallysalaglylurthrvalthrPheylsalaglylysasnleuyls 466
697 TTGCACAAAATGATCAGGT..... 717
467 vallysglnaspllyalasnphetrtyrserleuglnasplaleuth 483
718 .....GGTGCACAGTCACTTATGAGTACGCAAA 745
483 rglyleuthrserllethrleuglyglythrThrAsnsllyglnaspsa 500
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
746 AA.....ATTAACATAGCCCATATGCTTTTTCACCAACAGAGGC 786
500 lalysThrValilleasnllyaspglyleuthrillethrProaladlyasn 516
787 TCATTTGGGACAGTGGCTCACCAATGTTTATCTATGATGCCAAGCA 836
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
517 Gly...GlythrThrlythrasnhrilleser..... 526
837 AAGTGTATTAATTAATGGGATTCGCAAAACAGCAACCCCTATATA...G 883
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
527 .....Valthrlyaspglylilleysalaglysnlysalalierhfa 541
884 GAAAGCAATAGCTTCCAGCTAGTTCGTAAGATTTGTTCTATGATGA 933
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
541 snvalalaserglyleuargala.....Tyraspsp 551
934 ATTTTGTGAGATACCATTCAGTATCTACGAACACATCAAAATGAG 983
552 Ala..... 552
984 GAAATACCTTTTAAACGACATATATATGCGCAGAAATAACGATGCCA 1033
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:
553 .....Asnpheasplavalleunasnserlathraspleu..... 564

```

1034 AACATAACACTATCTCTACCTTATAGATTAAAAACGACGACGCTTCA 1083
565 ..AsnArgHisValGlnAspAlaTyrLys.....Gly 574
1084 TTGTTTATGTTTCTTTATCCGAGACAGACGACGACCTGTTAT..... 1128
575 LeuLeuAsnLeuAsnGlnLysAsnAlaAsnLysGlnProLeuValThrAs 591
1129 .CATGCTGAGGTGGGTCAACAGTTATCCGCCAGACTGAATAAAGAG 1177
591 pSerThrAlaAlaThrValGlyAspLeuArgLys..... 602
1178 AAAATATTTCTTTATTGACAAAGAAAGGTGAATTACTTACCGAC 1227
603LeuGlyTyrPValValSerThrLys 610
1228 AACATCAACCAAGCGCGCGGCTTTGTATTTGAGGTAATTTTACGGT 1277
611 Asn.....GlyThrLysGlnGlnLysSerAsn...GlnVal 620
1278 CTCGCTCAAAAACCAAGCAAGGCGGCGGCTTCATATACAGTG 1327
620 LysGlnAlaAspGlnValLeuPheThrGlyAlaGly..... 632
1328 ATGGCAGTACCGTTACTTGAAGAAGTAAACGCGTGCACACGACCGCTG 1377
633 ..AlaAlaThrValThrSerLys..... 639
1378 TCCAAATTCGCAAGGACGCGTGCCTGCTTCAAGCCAAAGGGAACCA 1427
640 SerGlnAsnGlnLysHisThrLys..... 647
1428 AGGCTCGTACGCGTGGCGACGCTAAAGTCATCTTATGATCAGACGCGG 1477
648ThrValSerValAlaGlnThrLys..... 655
1478 ACGATCAAGGCAAAAACCAAGCTTTAGTAATGCGCTGTGACGCGC 1527
656AlaAspLysGlnLysGlnLysAsp 663
1528 AGGGGACGCTGCACTGAATGCCGATATTCAGTTCAACCCGACCAACT 1577
664 GlyAspThrLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680
1578 CTATTTCGCTTTCGCGCGGACGCTTGGATTGAACGCGCATTCGCTT 1627
680 uThrValGlyAsnAsnGlyThrAlaValThrLysGlnGlyPheGlnThr 697
1628 CGTTCACGCGCATTCAAATACCGATGAAGGCGCATGTTGTCAC... 1674
697 allYsThrGlnAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713
1675CACATCAAGCAAAAGA..... 1692
714 AlaThrAlaAsnAspAlaAspLysLysValAlaAlaThrValLysAspValAl 730
1693TCACCGCTTACCATTAACGCAATAAAGATATTACTA 1729
730 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGlnLysLeuThr 747
1730 CAACC...GGCAATACAACTGGATAGCAAAAAAAGAAATGCGCTAC 1776
747 hSerLysLeuAspLysAsnProThrAspAsnGlyLysAspAlaLeu 763
1777 AACGCTGG.....TTTGCAGAAAGATGACACCA 1808
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVal 780
1809 AACGACGCGCG.....CTCAATCTGAATTACC 1837
780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGlnVal 797

1838 AACGGGAAGACGAGATCGCACTTACTGCTTCCGCGGACAAATTTA 1887
797 yThrAlaLysValSerAspThrLeuThrLysGlnAsnThrProThr 813
1888 AACGCG.....AATATACGCAAAACAAAGCGCA 1916
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspG 830
1917 ACTGTTTTCAGCGGACGACCGACCGCGCTCAATCATTTAGGA 1966
830 yLeuAsnPheAla...LysGlnThrAlaAspAla.....SerGlyS 843
1967 GCGGAGTCAAAAATGAAAGTATCCACAGACAGAAATCGTGGGAC 2016
843 erLysAsnValTyrLeuLysGlyIleAla..... 852
2017 AACGATTGATCGACCGCATTTAAACGGGAACCTTCATATTCAG... 2064
853 ThrThrLeuThrGlnProSerAlaGlyAlaLysSerSerHisValAspLe 869
2065 .GCGGACAAACGCGGTTCGCCAATGTTCGCAAGTGAA..... 2106
869 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGlnAspValL 886
2107GGCGATTGCAATTTAGCAATCAGCCCAAGCAATTTGCGGTGC 2151
886 euArgAlaGlyTyrPheAsnIleGlnLysAsnValAspTyrVal 902
2152 GCACCGCATCAAAAGCCACAACTCTGACAGCTTGACGTGACGGGCT 2201
903 AlaThrTyrAsp....ThrValAsnPheThrAspSerThrGlyTh 917
2202 GACAAGTTTACCGCAAAAACCATTCAGCAGATAA.....GTGATG 2245
917 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspValL 934
2246 CTTCATTGAGCAAGACCGACATC..... 2268
934 ystIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950
2269AGAGCAATGTGACCTTGGCGCA 2291
951 ThrGlyLysAspLeuLysAspAlaAsnGlnAlaThrValSerGlnAs 967
2292 TCACGCTCTTAAATCTACAGGACTTCCACACTCAGCGCACTTTA 2341
967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 984
2342 GTGACGCGGAGACACGCACTATACGCTTAGC...CGCAACGCGCACCA 2388
984 spAlaValAsnLysSerGlyTyrPheValThrGlyGlnGlyAlaThrAla 1000
2389 AACGCGAACCTCAGCCTCGTG.....GGCAATGCCCAAGCAATTTAA 2432
1001 GlnThrGlyAlaThrAlaValAlaAsnAlaGlyAsnAlaGlnThrVal 1017
2433 TCACGCACTTA.....AACGCGAAC.....ACATCGG 2461
1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1034
2462 CTTCGACAAATGCTTCATTTATCTAAGCAACAGCGCGTACAAAGCG 2511
1034 erLysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyAsp 1050
2512 AGCTGACGCTTTCGACACACGCTTAAGCAAAACGTAACGATTCGCACT 2561
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrIle 1067
2562 C.....AACGCAATGTCTCCCTA..... 2580
1067 uThrValThrGlyLysValSerValProAlaGlyAlaAsnSerValA 1084
2581GCCGATAAGCAGTATTCATTGAA 2607


```
1084 snasnasnLysLysLeuValasnaLaglInglLeuAlaThrAlaLeuAsn 1100
2608 AACAGCCGCTTACCGGAAAA..... 2628
1101 AsnLeuSerThrThrAlaLysAlaAspLysThrAlaAspGlyIuSerGI 1117
2629 .....ATCAGCGCGCG..... 2640
1117 uGIygluThrAspGlnIuValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AAGGATACGGCATTCAC 2658
1134 lagLysAsnLeuLysValLysIuSerGlyLysAspPheThrTyrSer 1150
2659 TTTAAAGACAGCAATGACGCGTCCGTCG.....GGCAGCGA 2696
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyIyThrAl 1167
2697 ATTAGGCAT.....TTAACCCTTGACACGCCACCA 2728
1167 aasnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThr 1184
2729 TTACAGCTCAATTCGCTATGACACGATGCGGCGGCGCAAAACGCGC 2778
1184 IeThrLeu.....AlaasnGlyAlaAlaLagly 1193
2779 AGTGGCGAGATGCGCGCGCGCGCTTCGCGCGCTTATTCCT 2828
1194 ThrAspAla.....SerasnGlyAsnThrIleSerVa 1204
2829 TACGGCGCCAACTTCGGCAGATCCGTTTCAACAGCGTGAACGTAACG 2878
1204 IThr.....LysAspGlyIleSerAlaGlyAsnLysGluIeThrAsnV 1219
2879 GCAAAATGAACGTCAGGAGACATTCGCTTATGTCGAACTTCGCGC 2928
1219 alLys.....SerAlaLeuLysThr 1225
2929 TACCGCAGCGCAATTTGAAGCTGGCGGAAATTCGGAAGCACTTACAC 2978
1226 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheH 1242
2979 CTGGGCTGACACATATCCGGCAGC...GAACCGTAAGTCTGACCAAT 3025
1242 salAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1259
3026 TGACGGTACTGAAAGAAAAAGACACACACCGCTGTCGAAAAATCTTAAT 3075
1259 IeThrValSerAlaLysThrAspAsn..... 1267
3076 TTACACCTGCAAAAGACACGTCGATGCGCGGCGATGGGTTATCAGCT 3125
1267 ..... 1267
3126 TATCCGAAAGACGCGGAGTTCGCTGCATATCCGTCGAAGAACAAG 3175
1268 .....AsnGlyLysHisThrValThrIleAspValAlaGluAlaH 1281
3176 AGCTTTCGCAAACTCGCAGCGGCGGAAACAGACAGCGCGCTTGAGC 3225
1281 yValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1295
3226 GCAAAA.....CAGCGACA 3239
1296 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1312
3240 ACTTCCCGCAAAACACAGCGGAAAAAGACACGCGCAAAAGCTTCAGC 3289
1312 rLysGlyAlaSerValAlaLysGlyIuPheAsnAlaValThrThrAspA 1329
3290 CGCTGATGGCGCGCGCGCATGCGCAAGCAAAAGGCAAAAGCTTGGC 3339
1329 IeThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValVal 1345
3340 GAACGCGCGCGCGCAGCGGAGGAAATGCGCATTTATGACGCGGA 3389
1346 Lys.....GlySerAsn...GlyAlaThrAlaThrGcl 1355
3390 GGAAGAGAAAAACGGGTGACGCGGATTAAGACACCGCGCTTGGCAAAAC 3439
1355 uThrAspLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1372
3440 AGCGCGAAGCG.....GAACCGCGCGCTACACCGCGC 3474
1372 snAspAlaIeThrPheValLysValGluAsnAspSerIeAlaThrIle 1388
3475 TTCCCCCGCGCGCGCGCGCGCGGATTCGCCAAGCGAGCGCCA 3524
1389 .....AspAspSerProThrAspAspLysAl 1397
3525 ACCGCAACCCCAACCGCAGCGCATGATCAGCCGTTATGCCAATAGCG 3574
1397 aasnAspAlaLeuLysAlaLagLyspThrLeuThrLeuLysAlaGlyLysA 1414
3575 GTTTGAGTGAATTTCCGCCAGCGCTCACAAGCGTTTGGCGGACGAGC 3624
1414 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1430
3625 GAATTGAGCGCGGTGTTGGCGAAGACCGCGCGCATTTGGACAAAG 3674
1431 AspLeu.....SerAlaLysSerAl 1437
3675 CGGCATCCGCGACACCAACACTACGTTGCAAGATTTCCGCGCGCTAAC 3724
1437 aThrValSerAsp..... 1441
3725 GCCAACAACCGACCTGCGCCAAATCGTATGCAAAAAACCTGCGAGC 3774
1442 .....LysLeuSerLeuGlyThrAsnLysAsnLysValAsnIleThrSer 1456
3775 GGGCGCGTGGCATCTGTTTCG...CACACCGGACCGGA..... 3813
1457 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrIleAspAspAl 1473
3814 AACACCTTCAGACGACGATCGGCAACCTGCGCAGCGCTGCGCACGCTG 3862
1473 aasnIleHisLeuAsnGlyIleAlaSerIleThr..... 1485
3863 CGGTTTGGGCAATACGCGATCGCAGGTTCGACATCGGCATGAGCGG 3912
1486 .....AspThrLeuLeuAsnSer 1491
3913 GGGCGGGTTTTATGAGCGCACCTTTACAGCGCATCAGAGCAAAAT 3962
1492 GlyAlaThrThrAsnLeuGly.....GlyAsnGlyIleThrAspAsnGcl 1506
3963 CCGCGCGCGCGGTGTCATTAAGCATTCAGGCAAGATACCGCGAGGT 4012
1506 uLysLysArgAlaAla.....SerValLysAspValLeuAsnAlaGlyT 1521
4013 TCGCGGATTCGCGATTCGAACCGCACATCGGCGCAACGCGCTATTTCGTC 4062
1521 rPAsnValArgLysValLysPro..... 1528
4063 CAAAAAGCGGATTCAGATACGAAACGTCATATCGCACCGCGGCGCT 4112
1529 AlaSerAlaAsnAsnGluValGluAsnIleAspPheValAla..... 1542
4113 TGCATTCAACCGCTACCGCGCGGCGATTAGGCAAGATTATTCATTAAC 4162
1542 ..... 1542
4163 CGGCGCAACACTTTCATACGCGCTTATTTAGCGTGCCTATACCAT 4212
1543 .....ThrTyrAspThrValAsp 1548
```

```
4213 GCCGCTCCGCGAAGTCCGACGGCGTCATATACCGCCGTATTGGCGCA 4262
      ::::::::::::::::::::
1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerLys 1564
      ::::::::::::::::::::
4263 GGATTTCGCGAAACCAGCCAGTGGGATGGCGGTAAACCGCAATC. 4311
      ::::::::::::::::::::
1564 saspsasnlylsarthythrGluValLysIleGlyAlaLysThrSerVal 1581
      ::::::::::::::::::::
4312 .....AAAGTTTCACGCTGTCCCTCCACGGCGCCGCC 4344
      ::::::::::::::::::::
1581 IelYsAspHisAsnGlyLysLeuSerThr..... 1590
      ::::::::::::::::::::
4345 GCCAAGGGCGCCCAATTGGAAAGCCACACAGCGCGGCGATC 4386
      ::::::::::::::::::::
1591 .....GlyLysGluLeuLysAspAlaAsnAsnAsnGlyVal 1602
      ::::::::::::::::::::

seq_name: /cogn2_6/protodata/1/1aa/6B_COMB.pep:US-09-669-974-33

seq_documentation_block:
/ Sequence 33, Application US/09669974
/ Patent No. 6333173
/ GENERAL INFORMATION:
/ APPLICANT: PEAK, Ian Richard Anselm
/ APPLICANT: MOXON, E. Richard
/ APPLICANT: JENNINGS, Michael Paul
/ TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
/ FILE REFERENCE: 065064/0128
/ CURRENT APPLICATION NUMBER: US/09/669,974
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US 09/377,155
/ PRIOR FILING DATE: 1999-08-19
/ PRIOR APPLICATION NUMBER: PCT/AU98/01031
/ PRIOR FILING DATE: 1998-12-14
/ PRIOR APPLICATION NUMBER: GB 9726398.2
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 33
/ LENGTH: 2353
/ TYPE: PRT
/ ORGANISM: Haemophilus influenzae
US-09-669-974-33

alignment_scores:
      Quality: 253.50      Length: 1664
      Ratio: 0.344      Gaps: 75
      Percent Similarity: 44.231      Percent Identity: 19.591

alignment_block:
US-09-303-518D-653 x US-09-669-974-33 ..
Align seg 1/1 to: US-09-669-974-33 from: 1 to: 2353

7 AGAACCGCAACGACCAACCGAA.....ACACACCGCAAAAGCCCC 47
  ||| ||| ||| ||| ||| ::| ||| ::| |||
246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnG 262
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
48 TAAACCGCGCGCATCCGCTTCGCGCGCTACTAGCCATATGCGCTGT 97
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 YlySerThrThrGluValLysPheThrProLysThrSerValIleLysG 279
  ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
98 CGTTCGGCAATTCCTGCCCAAGCCGCGGCGGACACACTTATTTCGCGATC 147
  ::| ::| ||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
279 ysaspglYlyLeuSerThrGlyLysGluAsnAsnAspThrAsnLysVal 295
  ::| ::| ||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
148 AACTACCAATACTATCGGACTTTGGCGAAATAAAGCAAGTTGGCACT 197
  ::| ::| ||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
296 ThrSerAsnThrAlaThrAspAsnThrAspGluLysAsnGlyLeuValTh 312
  ::| ::| ||| ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
198 CGGGCGCAAGATATTTGAGTTTACACCAAAAGGCAAGTTGCTGGCA 247
  ::| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


1084 snansnlnslslyslsleuValasnlaclncllyleuAlatThrAlaLeuasn 1100
2608 AACAGCCGCTTACCGGAA... 2628
1101 AsnLeuSerTrpThrAlaIysAlaAspLysTyrAlaAspGlyIuSerGI 1117
2629ATCAGCGCGCG... 2640
1117 uclglutThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134
2641AAGGATACGCAATTAC 2658
1134 lclgllyslnleuLysValLysGlnSerGlyLysAspPheThrTyrSer 1150
2659 TTAAGAAGACAGCAATGAGCGTCGCTG...GGCAGCA 2696
1151 LeuGlnAspThrLeuThrGlyLeuThrSerLeuThrLeuGlyLysThrAl 1167
2697 ATTAGGCAT...TTAAACCTTGACAAAGCCACCA 2728
1167 aasncllyAspAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1184
2729 TTACACTGAATTCGCGCTATGACAGACGATGCGAGCGCGGCAACGCGC 2778
1184 lcthrleu...AlaAsnclYAlaIlaAlaGly 1193
2779 AGTGGCGAGATGCGCGCGCGCGCTGCGCGCGCTTCCATATATCCGT 2828
1194 ThrAla...SerAsnclYAsnThrIleSerVa 1204
2829 TACGCGCGCACTTGGGAGATCCGTTGCAACACGCTGAGCGTAACG 2878
1204 lThr...LysAspclYIleSerAlaGlyAsnLysGluIleThrAsnV 1219
2879 GCAATTGACGCTGACGAGCAATCCGCTTATGTCGCACTTCTCGC 2928
1219 allys...SerAlaLeuLysThr 1225
2929 TACGCGAGCGCAATTGAACTGCGGGAATGTCGGAAGCACTTAC 2978
1226 TyrLysAspThrGlnAsnThrAlaAspGlyThrGlnAspLysGluPhe 1242
2979 CTGGCTGTCAACATACCGGCAAC...GAAACCTGAAGTCTCGAGCAAT 3025
1242 salAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1259
3026 TGAGCGTATGAGGAAAGAACACACACGCGCTTCCGAAATCTTAAT 3075
1259 latThrValSerAlaLysThrAspAsn... 1267
3076 TTCACCTGCAAAAGCAACGTCGATGCGCGCGCATGGCTTATACGT 3125
1267 1267
3126 TATCCGCAAAAGCGGAGTTCGCGCTGCAATTCGCTCAAGAACAG 3175
1268AsnclYLeuAsnIleThrValThrIleAspValAlaGluLysAla 1281
3176 AGCTTCCGCAACATCGGCAAGCGGGAAGAAACAGAGCGCGCTTACG 3225
1281 ysalGlyAspGlyLeuGluLys...AspThrAspGlyLysIleLys 1295
3226 GCAAA...CAGGACCA 3239
1296 leuLysValAspAsnThrAspLysAsnlnleuThrValAspAlaThr 1312
3240 ACTTGGCGGCAACAGCGGCGGAAAGACACGCGCAAGCGCTTACG 3289
1312 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAsp 1329
3290 CGCTGATTCGGCGCGGCAATGCCACCGAAAGGACAGAAATGTTGCC 3339

seq_name: /cgn2_6/prodata/1/laa/5A.COMB.pep:US-08-409-995-4

seq_documentation_block:

Sequence 4, Application US/08409995

Patent No. 5646259

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen I.

APPLICANT: St. Gene III, Joseph W.

TITLE OF INVENTION: Haemophilus Adhesion Proteins

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/409,995

FILING DATE: 24-MAR-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-61053/RFT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1912 amino acids

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: unknown

US-08-409-995-4

alignment_scores:

Quality: 248.50 Length: 1577

Ratio: 0.350 Gaps: 70

Percent Similarity: 44.959 Percent Identity: 19.784

alignment_block:

US-09-303-518D-653 x US-08-409-995-4 ..

Align seq 1/1 to: US-08-409-995-4 from: 1 to: 1912

7 ACAACGCAAAAGCGACACGCAAA.....ACACACGCAAAAGCCG 47

|||||

245 ThrilyspLysasnThrLeuAspValValLeuThrAlaLysGluasn** 261

48 TAAACGCGCCGCAATCGAATGGAGCGCACACAGCGCGATTTATTCGCGAT 278

|||||

261 LysThrThrGluValLysPheThrProLysThrSerValLleLysGlu 278

98 CGTTGCGGATTCGCGCCGCAAGCGCGCGCGGACACATTTATTCGCGAT 147

|||||

98 CGTTGCGGATTCGCGCCGCAAGCGCGCGCGGACACATTTATTCGCGAT 147

148 AACTACCAATCTATCGGACTTTCGCGAAATAAGCAATTTGCACT 197

|||||

148 AACTACCAATCTATCGGACTTTCGCGAAATAAGCAATTTGCACT 197

278 YAspGlyLysLeuPheThrLysGluAsnAsnAspThrAsnLysVal 294

|||||

278 YAspGlyLysLeuPheThrLysGluAsnAsnAspThrAsnLysVal 294

295 ThrSerAsnThrAlaThrAspAsnThrAspGluLysAsnLysValTh 311

|||||

295 ThrSerAsnThrAlaThrAspAsnThrAspGluLysAsnLysValTh 311

198 CGGCGCGAAAGATATTGAGGTTTACACAAAAAAGGAGTGGTCGCGCA 247

|||||

198 CGGCGCGAAAGATATTGAGGTTTACACAAAAAAGGAGTGGTCGCGCA 247

311 rAlaLysAlaValLleAspAlaValAsnLysAlaGlyTrpArgLysT 328

|||||

311 rAlaLysAlaValLleAspAlaValAsnLysAlaGlyTrpArgLysT 328

248 AATCGATGCGAAGCGCGATGATTTT.....TCGTGTATTCGCGT 294

|||||

248 AATCGATGCGAAGCGCGATGATTTT.....TCGTGTATTCGCGT 294

328 hThrThrAlaAsnGluAsnGlyAspPheAlaThrValAlaSerGly 344

|||||

328 hThrThrAlaAsnGluAsnGlyAspPheAlaThrValAlaSerGly 344

295 AACGCGGTGGCGCATTTGGCGGCGATCAATATTGTCGCGCGCA 344

|||||

295 AACGCGGTGGCGCATTTGGCGGCGATCAATATTGTCGCGCGCA 344

345 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 361

|||||

345 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 361

345 TAAAGCGCGCTTAAACAATTTGATTTTGTGCGGAGGAAGCAATCCG 394

|||||

345 TAAAGCGCGCTTAAACAATTTGATTTTGTGCGGAGGAAGCAATCCG 394

361 sAspThrAsnGlyAsnGly..... 367

|||||

361 sAspThrAsnGlyAsnGly..... 367

395 ATAGACACGCGTTTCTTACCAAAATTTGAAAAAATAATTTAAAGCA 444

|||||

395 ATAGACACGCGTTTCTTACCAAAATTTGAAAAAATAATTTAAAGCA 444

368LleThrValLysTyrAspAlaLysVal 376

|||||

368LleThrValLysTyrAspAlaLysVal 376

445 GGGACTAAGCGCCATCTTATGCGCGCATTTATATGCCGCTTTGA 494

|||||

445 GGGACTAAGCGCCATCTTATGCGCGCATTTATATGCCGCTTTGA 494

377 Gly...AspGlyLeuLysPheAspSerAsp.....Ly 386

|||||

377 Gly...AspGlyLeuLysPheAspSerAsp.....Ly 386

495 CAATTTGTCACGATGCGAAGCCTGTAGATGACCACTTATATGATG 544

|||||

495 CAATTTGTCACGATGCGAAGCCTGTAGATGACCACTTATATGATG 544

386 sLysLleValAlaAspThrThrAlaLeuThrValThr.....G 399

|||||

386 sLysLleValAlaAspThrThrAlaLeuThrValThr.....G 399

545 GGTGGAATACGCTGATTAATATATACCTGATGCTGTGCAATC... 591

|||||

545 GGTGGAATACGCTGATTAATATATACCTGATGCTGTGCAATC... 591

399 LysGlyLysValAlaGluLleAlaLysGluAspLysLysLysLeuVal 415

|||||

399 LysGlyLysValAlaGluLleAlaLysGluAspLysLysLysLeuVal 415

592GAGCAGCGCACATATTGGCGG.... 615

|||||

592GAGCAGCGCACATATTGGCGG.... 615

416 AsnAlaGlyAspLeuValThrAlaLysLysLysLysLysLysAlaLy 432

|||||

416 AsnAlaGlyAspLeuValThrAlaLysLysLysLysLysLysAlaLy 432

616TCGTATGCAAGC.....GAGCCCAATTAACCGGAAGTT 649

|||||

616TCGTATGCAAGC.....GAGCCCAATTAACCGGAAGTT 649

432 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyLleSerLysAspGln 449

|||||

432 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyLleSerLysAspGln 449

650 CATATCATATTGCAAGCGCATTTCTGCTGCTGCGTGGCAAT...ACC 696

|||||

650 CATATCATATTGCAAGCGCATTTCTGCTGCTGCGTGGCAAT...ACC 696

449 LuValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 465

|||||

449 LuValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuLys 465

697 TTTCACAAAAATGATCAGT..... 717

|||||

697 TTTCACAAAAATGATCAGT..... 717

466 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 482

|||||

466 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 482

718GGTGGCAGATGCACTTATGATGAGCGAA 745

|||||

718GGTGGCAGATGCACTTATGATGAGCGAA 745

482 rGlyLeuThrSerLleThrLeuGlyGlyThrThrAsnGlyLysAsnAsp 499

|||||

482 rGlyLeuThrSerLleThrLeuGlyGlyThrThrAsnGlyLysAsnAsp 499

746 AA.....ATTAAACATAGCCCATATGCTTTTATACACAGAGAGC 786

|||||

746 AA.....ATTAAACATAGCCCATATGCTTTTATACACAGAGAGC 786

499 lAlysThrValLleAsnLysAspGlyLeuThrLleThrProAlaGlyAsn 515

|||||

499 lAlysThrValLleAsnLysAspGlyLeuThrLleThrProAlaGlyAsn 515

787 TCATTTGGCGCAGAGTGTCAACATGTTATCTATGATGCCCAAAAGCA 836

|||||

787 TCATTTGGCGCAGAGTGTCAACATGTTATCTATGATGCCCAAAAGCA 836

516 Gly...GlyThrThrGlyThrAsnThrLleSer..... 525

|||||

516 Gly...GlyThrThrGlyThrAsnThrLleSer..... 525

```

837 AAGGCTTAATTAATGGGTATTGCAACAGCAACCCCTATATA...G 883
      ::::: ::::: ::::: ::::: :::::
526 .....ValThrLysAspGlyIleIysAlaGlyAsnLysAlaIleThra 540
      ::::: ::::: ::::: ::::: :::::
884 GAAAAAGCAATGGCTTCACGCTAGTTCGTAAGAATGGCTCTATGATGAA 933
      ::::: ::::: ::::: ::::: :::::
540 snValAlaSerGlyLeuAlaIleThraAla.....TyrAspAsp 550
      ::::: ::::: ::::: ::::: :::::
934 ATCTTGGCTGGAGATACCATTCAGTATTTACGACACATCAAAATGG 983
      ::::: ::::: ::::: ::::: :::::
551 Ala..... 551
      ::::: ::::: ::::: ::::: :::::
984 GAATACCTTTTAAACGCAATTAATATGCGCAGAAATGATGCCA 1033
      ::::: ::::: ::::: ::::: :::::
552 .....AsnIleAspValLeuAsnAsnSerAlaThrAspLeu..... 563
      ::::: ::::: ::::: ::::: :::::
1034 AACATTAACACTATTTCTACCTATAGATTTAAAAACACACACCGTTCAA 1083
      ::::: ::::: ::::: ::::: :::::
564 ..AsnATGHisValGluAspAlaThrLys.....Gly 573
      ::::: ::::: ::::: ::::: :::::
1084 TTGTTTAATGTTTCTTTATCCGACAGCAAGAACCTGTTAT..... 1128
      ::::: ::::: ::::: ::::: :::::
574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGluProLeuValThrAs 590
      ::::: ::::: ::::: ::::: :::::
1129 ..CATGCTGACAGGTGGGTCAACAGTTATTCACCCACACTGAATATGAG 1177
      ::::: ::::: ::::: ::::: :::::
590 pSerThrAlaIleThrValGluLysPleuArgLys..... 601
      ::::: ::::: ::::: ::::: :::::
1178 AAAATATTTCTTATTTAGCAAGAAAGTGAATGTACTTACCACG 1227
      ::::: ::::: ::::: ::::: :::::
602 .....LeuGlyTrpValValSerThrLys 609
      ::::: ::::: ::::: ::::: :::::
1228 AACATCAACAGCGCGGCGGCTTTGATTTTGAAGGTAATTTACGCT 1277
      ::::: ::::: ::::: ::::: :::::
610 Asn.....GlyThrLysGluGluSerAsn...GlnVal 619
      ::::: ::::: ::::: ::::: :::::
1278 CTGCGCTTAAACACAGCAACGCGGCGGCGGCTTATATCAGAG 1327
      ::::: ::::: ::::: ::::: :::::
619 LysGlnAlaAspGluValLeuPheThrIleAlaGly..... 631
      ::::: ::::: ::::: ::::: :::::
1328 ATGGCAGTACCGTTACTTGGAAAGTAAACGGCGTGCAACGACCGCTG 1377
      ::::: ::::: ::::: ::::: :::::
632 ..AlaIleThrValThrSerLys..... 638
      ::::: ::::: ::::: ::::: :::::
1378 TCCAAAATGCGCAAGAGCAGCGCTGCTTCAAGCAAGGGAACCA 1427
      ::::: ::::: ::::: ::::: :::::
639 SerGluAsnGlyLysHisThrIle..... 646
      ::::: ::::: ::::: ::::: :::::
1428 AGGCTGCGTACGGTGGCGAGCGGTAAAGTCATTTAGATCAGAGCGG 1477
      ::::: ::::: ::::: ::::: :::::
647 ....ThrValSerValAlaGluThrLys..... 654
      ::::: ::::: ::::: ::::: :::::
1478 AGCATGAAGCAAAAAACAGCCTTAGTGAATCGCGTGTGACGGCG 1527
      ::::: ::::: ::::: ::::: :::::
655 .....AlaAspCysGlyLeuGluLysAsp 662
      ::::: ::::: ::::: ::::: :::::
1528 AGGGGACGGTGCACATGATCCGATATCATGTCACCCGCAAACT 1577
      ::::: ::::: ::::: ::::: :::::
663 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 679
      ::::: ::::: ::::: ::::: :::::
1578 CTAATTCGGCTTTCGGCGCGGACGTTTGATTTGAAGCGGATTCGGCTT 1627
      ::::: ::::: ::::: ::::: :::::
679 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 696
      ::::: ::::: ::::: ::::: :::::
1628 CGTCCACCGCATTCAAAATACCGATGACGAGGCGATGATGTCAC... 1674
      ::::: ::::: ::::: ::::: :::::
696 alLysThrGlyAlaThrAspAlaAspArgLysValThrValLysAsp 712
      ::::: ::::: ::::: ::::: :::::
1675 .....CACAATCAAGCAAGAA..... 1692
      ::::: ::::: ::::: ::::: :::::
713 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 729
      ::::: ::::: ::::: ::::: :::::
1693 .....TTCACCGCTTACCATTTACAGGCAATTAAGATATTACTTA 1729
      ::::: ::::: ::::: ::::: :::::
729 aThrAlaIleAsnSerAlaIleThrPheValLysThrGluAsnLeuThrT 746
      ::::: ::::: ::::: ::::: :::::
1730 CAACC...GGCAATACACACACTTGATAGCAAAAAAGCAATTCCTAC 1776
      ::::: ::::: ::::: ::::: :::::
746 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAlaLeu 762
      ::::: ::::: ::::: ::::: :::::
1777 AACGGTTGG.....TTTGGCAGCAAGATGACCA 1808
      ::::: ::::: ::::: ::::: :::::
763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVally 779
      ::::: ::::: ::::: ::::: :::::
1809 AACGACGGCGCG.....CTCAATTCGATTTACC 1837
      ::::: ::::: ::::: ::::: :::::
779 sAlaAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluVal 796
      ::::: ::::: ::::: ::::: :::::
1838 AACCGGAGAAAGCGAGATCGCATTTACTGTTCCGCGGACCAAAATTA 1887
      ::::: ::::: ::::: ::::: :::::
796 ySThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 812
      ::::: ::::: ::::: ::::: :::::
1888 AACGGC.....AATATCCGCAAAACAGCGCA 1916
      ::::: ::::: ::::: ::::: :::::
813 GlyLysThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGl 829
      ::::: ::::: ::::: ::::: :::::
1917 ACTGTTTTCACGCGCGACACCGACCGACGCGCTTACATTTAGGAA 1966
      ::::: ::::: ::::: ::::: :::::
829 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyLys 842
      ::::: ::::: ::::: ::::: :::::
1967 GCGGCTGCTCAAAAATGAAGTATCCACAGAGAAATCGTGGGAC 2016
      ::::: ::::: ::::: ::::: :::::
842 erLysAsnValTyrLeuLysGlyIleAla..... 851
      ::::: ::::: ::::: ::::: :::::
2017 AACGATTTGATGACCGCACATTTAAACGGGAAACTTCATATTAC... 2064
      ::::: ::::: ::::: ::::: :::::
852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 868
      ::::: ::::: ::::: ::::: :::::
2065 ..GGCGGACAAAGCGGTGTTCCCGCAATGTTGCCAAATGGAA..... 2106
      ::::: ::::: ::::: ::::: :::::
868 uAsnValAspAlaThrLysLysSerAsnAlaIleSerIleGluAspVal 885
      ::::: ::::: ::::: ::::: :::::
2107 ....GGCATTTGGCATTTAAGCAATCACGCCCAAGCATTTTGGGTGC 2151
      ::::: ::::: ::::: ::::: :::::
885 euAlaGlyAlaLysThrAsnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 901
      ::::: ::::: ::::: ::::: :::::
2152 GCACCGCATCAAGCCACACATCTGTACAGTTGCGACGTGCGAGGCT 2201
      ::::: ::::: ::::: ::::: :::::
902 AlaThrTyrAsp.....ThrValAsnPheThrAspSerThrGlyTh 916
      ::::: ::::: ::::: ::::: :::::
2202 GACAACTGTACCGAAAAAACCATTACCGAGCATPAA.....GTGATTG 2245
      ::::: ::::: ::::: ::::: :::::
916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspVal 933
      ::::: ::::: ::::: ::::: :::::
2246 CTTCAATGAGCAAGACGACATC..... 2268
      ::::: ::::: ::::: ::::: :::::
933 yIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949
      ::::: ::::: ::::: ::::: :::::
2269 .....AGAGCAATGCTCAGCCTTGCGCA 2291
      ::::: ::::: ::::: ::::: :::::
950 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 966
      ::::: ::::: ::::: ::::: :::::
2292 TCACGCTCATTTAATCTCAGAGACTTGCACACTCAGCAAGGCAATCTA 2341
      ::::: ::::: ::::: ::::: :::::
966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 983
      ::::: ::::: ::::: ::::: :::::
2342 GTGACGGGCGGAGACAGCATATACGTTACG...CGCAACGCCACCCCA 2388
      ::::: ::::: ::::: ::::: :::::
983 spAlaValAsnLysSerGlyTyrParGValThrGlyGlnGlyAlaThrAla 999
      ::::: ::::: ::::: ::::: :::::
2389 AACGCAACCTCAGCCTCGTG.....GGCAATGCCCAACCAACATTTAA 2432
      ::::: ::::: ::::: ::::: :::::
1000 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1016
      ::::: ::::: ::::: ::::: :::::
2433 TCAAGCCACATTA.....AACGGCAAC.....ACATCGG 2461
      ::::: ::::: ::::: ::::: :::::

```

1016 rGlyThiSerValAsnPhelyAsnGlyAsnAlaThrThrAlaThrValS 1033
2462 CTTCCGACAAATGCTTATTAATCTAGCAACACCGCCGTACAAACGGC 2511
1033 eTlysAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyasp 1049
2512 AGCTGACGCTTTCCGACACGCTTAGCAACGTAACCATTCGGCACT 2561
1050 GlyLeuLysIleGlyAspAspLysIleValAlaAspThrThrThr 1066
2562 C.....AACGCAATGTCCTCA..... 2580
1066 uThrValThrGlyLysValSerValProAlaGlyAlaAsnSerValA 1083
1083 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1099
2608 AACAGCCGCTTACCGGAAA..... 2628
1100 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyLysGrl 1116
2629ATCAGCGCGCGC..... 2640
1116 uGlyLysThrAspGlnGluValLysAlaGlyLysAspLysValThrPhelysA 1133
2641AAGGATACGGCATTTAC 2658
1133 laGlyLysAsnLeuLysValLysGlnSerGlyLysAspPheThrTyrSer 1149
2659 TTTAAAGACAGCAATGACGCGCGCTCG.....GGCAGCA 2696
1150 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyLysThrAl 1166
2697 ATTAGCAAT.....TTAAACCTTGACAACGCCACCA 2728
1166 aaGnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
2729 TTTCACTCAATTCGGCTATCGACAGATGCGCGGCAAGCGCAACGGC 2778
1183 leThrLeu.....AlaAsnGlyAlaAlaAlaGly 1192
2779 AGTGGCGACATGCGCGCGCGCGCTTATGCGGAACCTTCGGC 2828
1193 ThrAspAla.....SerAsnGlyAsnThrIleSerVal 1203
2829 TACGCGCGCACTTCGGCAGATCCGCTTCAACACGCTGACGTAACG 2878
1203 LThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1218
2879 GCAATTTGAACGGTCAGGAAACATTCGCTTATGCGGAACCTTCGGC 2928
1218 allYs.....SerAlaLeuLysThr 1224
2929 TACCGCAGCGCAATTTGAAGCTGGCGAAAGTTCCGAGGCACTTACAC 2978
1225 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheN 1241
2979 CTTGGCTGTCAACAATACCGGCAAC...GAACCGCTAAGTCTGAGCAAT 3025
1241 saAlaIleValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1258
3026 TGACGGTGTGGAAGAAAGCAACACACGCGCTGTCGGAATCTTAAT 3075
1258 laThrValSerAlaLysThrAspAsn..... 1266
3076 TTCACCTGTCAAAAGCAACACGTCGATGCCGCGCATGCGCTTTCAGCT 3125
1266 1266
3126 TATCGCAAAAGACGCGGAGTTCCGCTGCATATCCGCTCAAAAGCAAG 3175
1267AsnGlyLysHisThrValThrIleAspValAlaGluAlaL 1280

3176 AGCTTCCGACAAACTCGCCAGACGCGGAGAAACAGAGCCGCTTGACG 3225
1280 ysaValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1294
3226 GCAAA.....CAGCACA 3239
1295 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1311
3240 ACTTGGCCGCCAAACACACGCGGCAAAAGCAACGCGCAAGCTTGACG 3289
1311 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspa 1328
3290 CGCTGATTCGGCGCGCGCATGCGCAACGCAAGCAAGCAAGTGTGCC 3339
1328 laThrThrAlaGlnGlyThrAsnAlaAsnGluThrGlyLysValAlaVal 1344
3340 GAACCGCGCGCGGACAGCGCGGGAATTCGGCACTTTCAGACGCGCA 3389
1345 Lys.....GlySerAsn...GlyAlaThrAlaThrGrl 1354
3390 GGAAGAGAAACAGGGTGCAGCGGATTAAGACACCGCCTTGCGCAAC 3439
1354 uThrAspLysLysValAlaThrValGlyAspValAlaLysAlaLle 1371
3440 AGCGCGAAGCG.....GAACCGCGCGCTTACCAACCGCC 3474
1371 snAspAlaIleThrPheValLysValGluAsnAspSerAlaThrIle 1387
3475 TTCCCGCGCGCGCGCGCGCGCGGATTTGGCGCAACCGCGCCCA 3524
1388AspAspSerProThrAspAspGlyAl 1396
3525 ACCGCAACCCCAACCGCAGCGCATGATCAGCGCTTATGCAATAGCG 3574
1396 aAsnAspAlaLeuLysAla***AspThrLeuThrLeuLysAlaGlyLysA 1413
3575 GTTTGAGTGAATTTCCGCAACGCTCAACACGCTTTCGCCGTACAGAC 3624
1413 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1429
3625 GAATTTGACCGCGCTTTCGCAAGACGCGCAACGCTTGGACAG 3674
1430 AspLeu.....SerValLysSerAl 1436
3675 CGCATTCGCGGACACCAACACTACCTTCGCAAGATTTCCGCGCTTACC 3724
1436 aThrValSerAsp..... 1440
3725 GCCAACAAACGCACTGCGCCAAATCGATGACAGAAACCTGCGCAGC 3774
1441LysLeuSerLeuGlyThrAsnGlyAsnLysValAlaAsnIleThrSer 1455
3775 GGGCGCGTGGCATCCTGTTTCG...CACACCGGACCGGA..... 3813
1456 AspThrLysGlyLysLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1472
3814AACCTTTCGACGACGCGCATTCGGA 3838
1472 aaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuAla 1489
3839 ACTTCGCGCAGCGCTTGGCCACGCGGTTCGCGGCAATACGCGATCGCG 3888
1489 snSerGly.....AlaThrThrAsnLeuGlyLysGlyIleThrAs 1503
3889 AGGTGACATCGCATCAGCGCGCGCGGCTTTAGTAGCGCAGCCT 3938
1503 paGnGluLysLys***AlaIleSerValLysAspValLeuAsnAlaGlyT 1520
3939 TTCAGACGCGCATCAGAGGCAAAATCCGCGCC..... 3970
1520 rpaAsnValArgGlyValLysProAlaSerAlaAsnAsnGlnValGluAsn 1536

```

3971 .....GCTGCTCATTAACGGCATTCAGGCA 3996
1537 lIeaSpheValAlaThrTyrAspThrValAspHeValSerGlyAspLy 1553
3997 AGATACCCGCGAGTTTCGGCGGATGCGATCGAACCGCAGTCGGCC 4046
1553 sAspThrSerValThrValGluSerLysAspAsnGlyLysArgThrG 1570
4047 AACGGCTATTTCGTCCAAAAGCGGATT 4075
1570 luValIleGlyAlaLysThrSerVal 1579
seq_name: /cgm2_6/ptodata/1/iaa/GA.COMB.pep:US-08-685-467-4

```

```

seq_documentation_block:
  Sequence 4, Application US/08685467
  Patent No. 6060059
  GENERAL INFORMATION:
  APPLICANT: St. Gene III, Joseph W.
  TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
  STREET: Four Embarcadero Center, Suite 3400
  CITY: San Francisco
  STATE: California
  COUNTRY: United States
  ZIP: 94111-4187
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/685,467
  FILING DATE: 22-JUL-1996
  CLASSIFICATION: 424
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/409,995
  FILING DATE: 24-MAR-1995
  ATTORNEY/AGENT INFORMATION:
  NAME: Silva, Robin M.
  REGISTRATION NUMBER: 38,304
  REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 781-1989
  TELEFAX: (415) 398-3249
  TELEX: 910 272299
  INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1912 amino acids
  TYPE: amino acid
  STRANDEDNESS: unknown
  TOPOLOGY: unknown
  MOLECULE TYPE: protein
  US-08-685-467-4

```

```

alignment_scores:
  Quality: 248.50      Length: 1577
  Ratio: 0.350        Gaps: 70
  Percent Similarity: 44.959   Percent Identity: 19.784

```

alignment_block:

US-09-303-518D-653 x US-08-685-467-4 ..

Align seg 1/1 to: US-08-685-467-4 from: 1 to: 1912

```

7 ACACGCGAAGCGACAAACGGA.....ACACGCGAAGCGCC 47
||| |||||:|||||:|:| |||:|||||:
245 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsn** 261

```

```

48 TAAACGGCGCGCAGTCCGCTTCTCGCCGCTTACTTAACCATATGCGCTG 97
||||| :|||||:|||||:|:| |||:|:|:|
261 *LysThrThrIleValLysPheThrProLysThrSerValIleLysGlu 278
98 GGTTCGCGATTCGCCCAACCGCGGAGACACATATTTCGCGCATC 147
:|:| | | | :|:|:|:|:|:|:|:|:|:|:|
278 yAspGlyLysLeuPheThrGlyLysGluAsnAspThrAsnLysVal 294
148 AACTATGCAATATCTATCGGACTTGGCGAAATTAAGCAAGTTTCAGT 197
:|:| :|:| :|:|:|:|:|:|:|:|:|:|:|
295 ThrSerAsnThrAlaThrAspAsnThrAspGluGlyAsnGlyLeuValh 311
198 CGGGGCGCAAGATATTGAGTTTACACAAAAGGAGGAGTGGTGCGCA 247
:|:| :|:|:|:|:|:|:|:|:|:|:|
311 rAlaLysAlaValIleAspAlaValAsnLysAlaGlyTParValLysT 328
248 AATGATGACGAAAGCCCGCATGATGATT..TCTGGGATGCGCT 294
:|:| :|:|:|:|:|:|:|:|:|:|:|
328 hrThrThrAlaAsnGlyLysAsnGlyAspPheAlaThrValAlaSerGly 344
295 AACGGCGTGGCGGATGGCGGCGCATCAATATATTGAGCGTGCGACA 344
:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
345 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrLy 361
345 TAAAGGGGCTATACAAATGTTGTTGTGGCGGAGGAGCAATCCCG 394
:|:| :|:|:|:|:|:|:|:|:|:|:|
361 sAspThrAsnGlyAsnGly..... 367
395 ATGAGCACCGCTTCTTCAACAATTGTGAAGAATAATTATTAAGCA 444
|||:|:|:|:|:|:|:|:|:|:|:|
368 .....IleThrValLysIleValAspAlaLysVal 376
445 GGAAGTAAAGCGCATCTTATGCGCGCATATATATATATATATATATAT 494
||| :|:|:|:|:|:|:|:|:|:|:|
377 Gly...AspGlyLeuLysPheAspSerAsp.....Ly 386
495 CAATTTGTGCAGAGATGCAAGACCTGTGACATGACCATATATATATG 544
|||||:|:|:|:|:|:|:|:|:|:|:|
386 sLysIleValAlaAspThrAlaLeuThrAlaLeuThrVal.....G 399
545 GGTGAATAATACCGTGAATTAATAATACCGCATGATGTTGCAATC... 591
|| ||| |||:|:|:|:|:|:|:|:|:|
399 yGlyLysValAlaGluIleAlaLysGluAspLysLysLeuVal 415
592 .....GGACGAGCAGACAAATATGCGG... 615
416 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerThrLysAlaLy 432
616 .....TCTGATGAAGAC.....GAACCAATATACCGCAAGTT 649
|||:|:|:|:|:|:|:|:|:|:|:|
432 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyLleSerLysAspGln 449
650 CATATCATATTGCAAGCGCATATTCTTGGTCGCGGCGCAAT...ACC 696
:|:| :|:|:|:|:|:|:|:|:|:|:|
449 luValLysAlaGlyLysIleThrValThrPheLysAlaGlyLysAsnLeuLys 465
697 TTTCGACAAATGATCAGT..... 717
|||:|:|:|:|:|:|:|:|:|:|:|
466 ValLysGlnAspGlyAlaAsnPheThrTyrSerLeuGlnAspAlaLeuTh 482
718 .....GGTGCACAGTCACTTATGTTAGTACGAA 745
|||:|:|:|:|:|:|:|:|:|:|:|
482 rLysThrThrSerIleThrLeuGlyGlyThrThrAsnGlyLysAsnAsp 499
746 AA.....ATTAAATAGCCCATATGTTTATTAACACAGGAGG 786
|||:|:|:|:|:|:|:|:|:|:|:|
499 lAlsThrValIleAsnLysAspGlyLeuThrIleThrProAlaGlyAsn 515
787 TCATTGGCGACAGTGGCTCACCAGTTTATCTATGATGCCCAAGCA 836
:|:| ||| :|:|:|:|:|:|:|:|:|:|:|
516 Gly...GlyThrThrGlyThrAsnThrLleSer..... 525
837 AAGTGTATATTAAATGGGATTTGCAACAGCAACCCCTATATA...G 883

```



```

526 .....ValThrLysAspGlyLeuLysAlaGlyAsnLysAlaLeuThra 540
540 .....ValThrLysAspGlyLeuLysAlaGlyAsnLysAlaLeuThra 540
884 GAAAGCAATGCGCTCCAGCTACTTGGTAAAGATGGTCTATGATGAA 933
540 snValAlaSerGlyLeuThrGAla.....TyrAspAsp 550
934 ATCTTGTGGAGATACCCATTCCAGTATTCTACGACACATCAAAATGG 983
551 Ala..... 551
984 GAATACTTTTAAACGACATTAATAGCGCGACGAAATTCGATGCA 1033
552 .....AsnPheAspValLeuAsnAsnSerAlaThrAspLeu..... 563
1034 AACATAACACTATCTCTACCTTATAGATTAAAAACACGACGCTTCAA 1083
564 ..AsnArgHisValGluAspAlaThrLys.....Gly 573
1084 TTGTTTAATGTTTCTTATCCGAGACACCAAGACACTGTTAT..... 1128
574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 590
1129 .CATGCTGAGGTGGGTCAACAGTTATGACCCAGACTGATATATGAG 1177
590 pSerThrAlaAlaThrValGlyAspLeuThrLys..... 601
1178 AAAATATTCTTATTGACAAAGGAAAAAGTGAATTGATCTATCCAGC 1227
602 .....LeuGlyTyrPValValSerThrLys 609
1228 AACATCAACCAAGCGCGCGCTTGTATTGAGGTAATTACGTT 1277
610 Asn.....GlyThrLysGluGluSerAsn...GlnAla 619
1278 CTGCGCTAAAAACACGAAACGTCGCGCGCTTCAATATCAGTG 1327
619 LysGlnAlaAspGluValLeuPheThrGlyAlaGly..... 631
1328 ATGGCAGTACCGTTACTTGAAGAATACGCGCGTGCACACCGCTG 1377
632 ..AlaAlaThrValThrSerLys..... 638
1378 TCCAAAATCGCAAGCAGCGCTGCTTCAAGCCAAAGGGAACA 1427
639 SerGluAsnGlyLysHisThrIle..... 646
1428 AGCGTCGTCAGCGTGGCGACGTAAGTCATCTTATGACGACGCG 1477
647 ....ThrValSerValAlaGluThrLys..... 654
1478 ACGATCAGGCAAAACAAAGCCTTATAGTAAATCGCGCTGTCACGCG 1527
655 .....AlaAspGlyLeuGluLysAsp 662
1528 AGGGGACGGTCACTGAATGCCGATATCACTTCAACCCGACAACT 1577
663 GlyAspThrLeuLysLeuLysValAspAsnGlnAsnThrAspAsnVal 679
1578 CTATTTGCGCTTTCGCGCGAGCTTGATTTGACGGCAGCTT 1627
679 uThrValGlyLysAsnGlyThrAlaValThrLysGlyGlyPheGluThr 696
1628 CGTTCACCGCATTCAAATACGATGAGGGGCGGATGTCAC... 1674
696 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 712
1675 .....CACATCAAGACAAAGAA..... 1692
713 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 729
1693 .....TTCACCGTTACCATTTACGGAATTAAGATTTTCTA 1729

```

```

729 arThrAlaLeuAsnSerAlaAlaThrPheValLysThrGluAsnLeuThr 746
1730 CAACC...GCCATTAACACCACTTGATATGCAAAAAAGAAATTCGCTAC 1776
746 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAlaLeu 762
1777 AACGGTTGG.....TTWGGCAGAAAGATGCAACCA 1808
763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVal 779
1809 AACGAACGGCGCG.....CTCATCTGATTTACC 1837
779 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluVal 796
1838 AACCGGAAGAGCGGATTCGACTTCTGCTTCCGCGGACAAATTTA 1887
796 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 812
1888 AACGCG.....AAATACGCAACAAAGCGCA 1916
813 GlyLysThrThrAlaThrProLysValAsnIleThrSerThrAlaAsp 829
1917 ACTGTTTTCAGCGCGACGACCGACCGCTCAATCATTTAGGA 1966
829 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 842
1967 GCGGTGTCTCAAAATGAAAGATCCACAGAGAAATGCTGTGGAC 2016
842 erLysAsnValTyrLeuLysGlyLeuAla..... 851
2017 AACGATTGATGACCGCACATTTAAAGCGAAACTTCCATTTG... 2064
852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAsp 868
2065 .GCGGACAAACGCGGTGTTCCGCAATGTCGCAAGTGAA..... 2106
868 uAsnValAspAlaThrLysLysSerAsnAlaIleSerIleGluAspVal 885
2107 .....GGCGATTGGCATTTAAGCAATCAGCCCAAGAGTTTGGTGTG 2151
885 euArgAlaGlyTyrAsnIleGlnGlyAsnGlyAsnValAspVal 901
2152 GCACCGCATCAAGCAGCACATCTGTACAGTTGCGAGTGGAGGCT 2201
902 AlaThrTyrAsp.....ThrValAsnPheThrAspSerThrGlyTh 916
2202 GACAGTGTGATCCGAAACCAATACCGACGATMAA.....GTGATG 2245
916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspVal 933
2246 CTTCATTGAGCAAGACCGACATC..... 2268
933 yeIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949
2269 .....AGAGCAATGTACGCTGCCGA 2291
950 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 966
2292 TCACGCTCATTTAAATCTCAGAGACTTCCGACACTCAACGGCAATCT 2341
966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 983
2342 GTGCAGCGCGACACGACGCTATGAGTTAGC...CGCAAGCCGACCCA 2388
983 spAlaValAlaLysSerGlyTyrPargValThrGlyGluLysAlaThrAla 999
2389 AACGGCAACCTGACCTCGTG.....GGCAATGCCAAGCAACATTTAA 2432
1000 GluThrGlyAlaThrAlaValAlaAsnAlaGlyAsnAlaGluThrValThr 1016
2433 TCAGGCAACATTA.....AACGGCAAC.....ACATCGG 2461
1016 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1033

```

2462 CTTCCGACAAATGCTTCATTATTAATCAACAAACGCCGTACAAAGCCG 2511
1033 eTlyAspAsnGlyAsnIleAsnValIyTyrAspValAsnValGlyAsp 1049
2512 ACTTGACGCTTTCCGACACGCTAAGCAACGTAAGCCATTCCGACCT 2561
1050 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrIle 1066
2562 C.....AACGGCAATGTCTCCCTA..... 2580
1066 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1083
2581GCCGATAAGGAGATTCATTTCATTGAA 2607
1083 snAsnAsnLysLysLeuValAsnAlaGlyLeuAlaThrAlaLeuAsn 1099
2608 AACAGCCGCTTTACCGGAAAA..... 2628
1100 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyLeuSerGI 1116
2629ATCAGCGCGCC..... 2640
1116 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1133
2641AAGGATACGGCATTCAC 2658
1133 IaGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSer 1149
2659 TTAAAGACAGCGAATGAGACGCTGCCGTCG.....GCGACGGA 2696
1150 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1166
2697 ATTAGGCAAT.....TTAAACCTTGACAAAGCCACCA 2728
1166 aAsnGlyIaIAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
2729 TTACACTCAATTCGCTATCGACAGATGCGGACGCGGCAAAACCGCG 2778
1183 LeThrLeu.....AlaAsnGlyAlaAlaAlaGly 1192
2779 AGTGGCGGACATGCGCGCGCGCGCTTGCGCGCTTCCATTATTCGCT 2828
1193 ThrAspAla.....SerAsnGlyAsnThrIleSerVal 1203
2829 TAGCGCGCAACTTCGCGAGAAATCCGTTTCAACACGCTGACGGTAACG 2878
1203 LThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1218
2879 GCAAATTTGAAGGTCAGGAGACATTCGCTTATGTCGGAACCTTTCGCG 2928
1218 allys.....SerAlaLeuLysThr 1224
2929 TTCCGAGGCGCAATTTGAGTGGCGGAAGTCCGAGGCACTTACAC 2978
1225 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheH 1241
2979 CTTGGCTGTCAACAATACCGGCAAC...GAACCGCTAGTCTCAGACAT 3025
1241 sAlaIaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1258
3026 TGACGGTACTGGAAGCAAAAGACACACGCTGTCCGAAATCTTAAT 3075
1258 IaThrValSerAlaLysThrAspAsn..... 1266
3076 TTCACCCGTGCAAAAGACACGTCGATGCCGCGCATAGCCTTATCAGCT 3125
1266 1266
3126 TATCCGAAAGACGCGAGATTCGCGCTGCATATCCGTCGAAGAACAAG 3175
1267AsnGlyLysHisThrValThrIleAspValAlaGluAla 1280

3176 AGCTTTCGCAAAATCTCGCAGCGGGAGAAACAGACAGCCGCTTGACG 3225
1280 ySValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1294
3226 GCAAAA.....CAGCGACA 3239
1295 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspIaThr 1311
3240 ACTTGGCGCCCAACACAGCGGCAAAAGACACGCGCAACGCTTGACG 3289
1311 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspA 1328
3290 CGCTGAATTCGCGCGCGGCGCATGCGCACGGAAGAACAGAACTGTTCG 3339
1328 IaThrThrAlaGlnGlyThrAsnAlaAsnGluThrGlyLysValValVal 1344
3340 GAACCGCGCGCGCGCAGCGAGCGGAAATCCGCGATTATGACGAGCGGA 3389
1345 Lys.....GlySerAsn...GlyAlaThrAlaThrGI 1354
3390 GGAAGAAAAAAGCGGTGCGAGCGGATTAAGACACCGCGCTTGCGCAAC 3439
1354 uThrAspLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1371
3440 AGCGCGAAGCG.....GAAACCGCGCGCTACACACCGCC 3474
1371 snAspAlaIaThrThrPheValLysValGluAsnAspSerIaThrIle 1387
3475 TTCCCGCGCGCGCGCGCGCGCGGATTTGCGGCAACGCGACCCCA 3524
1388AspAspSerProThrAspAspGlyAl 1396
3525 ACCGCAACCCCAACCGCAGCGGACCTGATCAGCGGTATGCGCATACG 3574
1396 aAsnAspAlaLeuLysAla**AspThrLeuThrLeuLysAlaGlyLysA 1413
3575 GTTTGAGTGATTTTCGCGCAGCTCAACAGCGTTTTCGCGTACAGAC 3624
1413 snLeuLysValLysArGAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1429
3625 GAATGGACCGCGTGTTCGCGAAGACCGCGGCAACGCGCTTTGACAAAG 3674
1430 AspLeu.....SerValLysSerAl 1436
3675 CGGCATCCGCGACACCAACTACCTGTCGAAGATTTCCGCGCATAC 3724
1436 aThrValSerAsp..... 1440
3725 GCCAACAACCGACTGCGCCCAATCGGTATGACGAAAAACCTGCGAGC 3774
1441LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1455
3775 GCGCGCGTGCAGCATCTGTTTCG...CACACCGGACCGCA..... 3813
1456 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1472
3814AACACCTTGACGACGATCGCA 3838
1472 aAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuLeuA 1489
3839 ACTGCGACAGGCTTGCGCAGCGTCCGCTTTGCGGCAATTCGCGATCGC 3888
1489 snSerGly.....AlaThrThrAsnLeuGlyGlyAsnGlyIleThr 1503
3889 AGTTTCGACATCGCATCGACGAGCGGGCGGTTTGATGTCGCGGCGGCT 3938
1503 pAsnGluLysLys**AlaAlaSerValLysAspValLeuAsnAlaGlyT 1520
3939 TTCAGACGCGATCAGAGGCAAAATCCGCGCG..... 3970
1520 rPAsnValArGlyValLysProAlaSerAlaAsnAsnGluValGluAsn 1536
3971GCGTGCTGCATTACGGCATTCAGGA 3996

```

1537 ILASpHeValaIaThrValrAspThrValaSPheValserLysApIy 1553
3997 AGATACCGCGCAGCTTTCGGCGATTTGGCATGACCGACATTCGGCGC 4046
1553 SASpThrThrservalThrValGIuSerLysASpANGlyLysArgThrg 1570
4047 AACGCGTATTTCTGCCAAAAGCGGATT 4075
1570 LuValLysIleGlyValaLysThrserval 1579

seq_name: /cgn2.6/ptodata/1/aa/6B_COMB.pep:US-09-413-814-78

seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hotte, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; TITLE OF INVENTION: heteropolypeptide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

alignment_scores:
Quality: 245.00 Length: 844
Ratio: 0.714 Gaps: 42
Percent Similarity: 40.640 Percent Identity: 24.526

alignment_block:
US-09-303-518D-653 x US-09-413-814-78 ..

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882

2539 GCAACGTAGCCATTCC.....GCACTCAACGG 2567
||||:|||||
75 AlameIleGlyHisSerLeuGIuLeuValaIaIaGlyeValaIaGI 91
2568 CAATGTCTCCCTAGCCGATAGACGACTATTCATTTTGAAACAGCCGCT 2617
||||| |||
91 yValPheThrLeuGIuAspAlaMetSerLeuValIleAspArgGIyAlrGp 108
2618 TTACCGGAAAAATCAGCGCGGAGAGATATCGCATTTACACTTA..... 2661
||||:||||: |||
108 hemeTGIaIaIaIaProIaGIySerMetLeuValaIaGIyLeuProIa 124
2662 AAAGACGAGATGACGCTGCCCTCGCGCAGCAATAGCAATTAA 2711
|||| ||| |||:||||| |||:||||
125 AlaSpValaGIuGIyLeuGIuValaGIyLeuGIuIleAlaIaValas 141
2712 CCTTGACAGCGCCACATTTACACTCAATTCCG..... 2743
|:||||: |||
141 nSerProLysLeuThrValaIaValaGIy.ProIaSerAlaIleArgAsp 157

```

2744 CCAATGACAC 2754
158 LeuAlaAlaArgLeuGluAlaArgGluValPheAlaArgProLeuGlnTh 174
2755 GATGGCGCG... GCGCGCAACCGCGAC 2780
174 TserHisAlaPheHisCysAlaLeuLeuAspGlyAlaValAlaProPheL 191
2781 TGGCGAGATGGCGCGCGCGCGCTGGCGCG... TTCCTATTAT 2824
191 encluserValArgArgAlaArgLeuSerProGluLeuProValVa 207
2825 C..... C 2826
207 IserAsnValThrGlyAlaLeuLeuThrAspAlaGluAlaThrAspPro 224
2827 GTTACGCCGCCAACTTGGCGAGATCCGTTTCAACACGCTGACGGTAA 2876
224 IerThrTrpAlaArgHisLeuArgGlnProValArgPheSerAspGlyVal 240
2877 CGGCAATTTGAACGGTCAAGAAC... ATTCCGTTTATTCGGAAC... 2921
241 GluAlaLeuPheAlaSerGlyHisAlaLeuPheLeuGluValGlyProG 257
2922CTTCGGCTTACCGCACCGCAATTGAAGCTGGCGGAAGTTCC 2964
257 YArgGlyLeuThrThrLeuValArgGlnThrLeuAlaGlyArgGlyGly 273
2965 GAAGGACATTACACCTTGGCTGTCAACATACCGGCAACGAACCCGGTAA 3014
274 AlaAlaLeuAlaSerLeuGlySerThrHisAlaAlaSerGluProAlaSe 290
3015 TTTCGA..... GCAATTGAC 3030
290 TruAlaGluAlaLeuGlyGlnLeuThrGlnAlaGlyHisAlaValAsp 307
3031 GTACTGGAGAAAGACAACACACCGCTGCCAAATCTTAATTTCAC 3080
307 rPrHisAlaArgProArgGlyArgProProAlaArgSerSer..... 320
3081 CCTGCAACGACACGCTGA..... TCGCGGCGCAT 3112
321 AlaGlySerArgThrArgSerSerGlyArgGlyThrGlySerThrArgAl 337
3113 GGCCTTATTCAGCTTATCCGCAAGACGGCGAGTCCGCTGCATATATCCG 3162
337 alaAlaAlaProProProProSerThrArgArg... ProThrArgProSerA 353
3163 GTCAAGACACAGAGCTTTCGCAACAATCGCGCAACGGCGGGAACACGA 3212
353 rGProArgArgThrProSerArgArgArgArgSerProAlaArgThrPro 369
3213 GCGCGCGCTTCGCGCAACAGGCAACACTGGCG..... 3247
370 Gly...ProArgSerProArgArgArgProSerProSerThrLargSerPro 385
3248CCAAACACAGCGCGCAAAAAACG 3270
366 AlaIleTrpIuArgProProArgArgProArgAspAlaArgProArg 402
3271 AACGGCGAAACCTTGAACGGCGTGAATTGGCG... CGGGCGCAATGC 3314
402 gLeuLeuArgProAlaArgArgLeuAlaArgAlaArgAspProGlyAspValP 419
419 roAspProArg.....AlaAlaArgArgArgAlaProAlaAlaGly 432
3315 CACGCAAAAGCGCAAGAGTGTTCGCAACCGCGCGCGGACGAGCGGGG 3364
419 roAspProArg.....AlaAlaArgArgArgAlaProAlaAlaGly 432
3365 AAAATCCGCAATTATGACGGCGGAGACAAGAAAAAGGGGCGACGGG 3412
433 AlaLeuProGlyAlaAspGlyArgGlyAlaArgGlyProGlyArgArgArg 449

```

3415 GATTAAGACACCGCTTGGCGAAGACGCGGAAACCGCGCGC 3464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 g.....AlaProArgArgGlyAlaIaIa.....AlaArgV 459
      : : : : : : : : : : : : : : : : : : : :
3465 TACGACCGCTTCCCGCGCGCGCGCGCGCGCGGATTGCCGCAAC 3514
      : : : : : : : : : : : : : : : : : : : :
459 aLrProArgArgArgAlaAspProArgArgProAlaProAlaVala 475
      : : : : : : : : : : : : : : : : : : : :
3515 C...GCACGCCCAACCGCAACCGCAACCGCGCGCGCGCGCGT 3561
      : : : : : : : : : : : : : : : : : : : :
475 lArgAlaIaIaProAlaValaIaArgArgSerAlaArgAlaGlyInPro. 491
      : : : : : : : : : : : : : : : : : : : :
3562 TATGCCAATACGCGTTGAGTGAATTTCCGCAACGCGCTTTT 3611
      : : : : : : : : : : : : : : : : : : : :
492 .....GlyValGlnHisPro 496
      : : : : : : : : : : : : : : : : : : : :
3612 CGCGCTACGAGCAATTTGACG...CGTGTTCGCGAGACCGCGCGCA 3658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 ValArgGlyAlaArgAspGlyProAlaArgArgArgArgAlaProG 513
      : : : : : : : : : : : : : : : : : : : :
3659 ACGCCGTTTGACACGCGCATCCG..... 3683
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513 nProValArgAspArgAlaIaIaProArgGlyAlaIaIaAspAspValGlna 530
      : : : : : : : : : : : : : : : : : : : :
3684 .....GGACACCAACACTACCGCTTCGCAAGATT..... 3713
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 rGlnGlnGlyArgArgAlaValProGlyArgArgAlaArgGlyProGlyAla 546
      : : : : : : : : : : : : : : : : : : : :
3714 .....CGCGCGCTACG..... 3725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
547 ValProAspValaArgProArgAlaProArgGlyArgGlyProArgArg 563
      : : : : : : : : : : : : : : : : : : : :
3726 .....CCACAA 3732
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
563 yGlyLeuGlyAlaArgProArgGlyLeuProArgAlaValaArgProGln 580
      : : : : : : : : : : : : : : : : : : : :
3733 ACCGACCTGGCCCAATCG.....TATGCAAAACCTGGGAG... 3773
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 roArgAlaIaIaProProArgAlaGlyHisProGlnAlaProArgArgAla 596
      : : : : : : : : : : : : : : : : : : : :
3774 .....CGGCGCGCTGCATCT.....GTTTTCGCA 3802
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 ArgHisArgAlaIaArgArgAlaArgValaArgArgLeuValGlyArg 613
      : : : : : : : : : : : : : : : : : : : :
3803 ACCGACCGCAACACCTTCGACAGCGGATCGCAACG.....GGCA 3846
      : : : : : : : : : : : : : : : : : : : :
613 gArgLeuArgArgArgAlaArgArgAlaLeuArgArgLeuArgAlaGlyP 630
      : : : : : : : : : : : : : : : : : : : :
3847 CGGCTTGCCGACGCGTGC.....CGTTTCGGCAATACGCGAT 3884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
630 roAlaIaIaProAlaIaGlyAlaIaProGlyAlaValaArgArgLeuArgArg 646
      : : : : : : : : : : : : : : : : : : : :
3885 CGGACGTTGCAATCGCATGACGCGCGCGCGGTTTGAAGCGCA 3934
      : : : : : : : : : : : : : : : : : : : :
647 SerProAlaGlyAlaValaValaArgArgGlyProGlyGly.....ArgAl 661
      : : : : : : : : : : : : : : : : : : : :
3935 GCGCTTC..... 3941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 aProValaIaIaAspArgAlaIaLeuGlyArgAlaIaIaProGlyProArg 678
      : : : : : : : : : : : : : : : : : : : :
3941 ..... 3941
      : : : : : : : : : : : : : : : : : : : :
678 lAgGlyArgProArgAlaGlyAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 694
      : : : : : : : : : : : : : : : : : : : :
3942 .....AGACGCGATCAGAGGAAATCCGCGCGCGCTGCATTACG 3985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
695 GluProArgArgArgGlyArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIa 711
      : : : : : : : : : : : : : : : : : : : :
3986 GCATTTCAGCAGATACCGCGCAG.....TTTCGCGGATTCGGCATC 4029
      : : : : : : : : : : : : : : : : : : : :
711 gAlaGlyArgLeuProLeuHisGlyProAlaIaIaIaIaIaIaIaIaIaIa 728
      : : : : : : : : : : : : : : : : : : : :
4030 GAACGCGA.....CATCGCGC 4046

```

```

seq_name: /cgn2_6/ptodate/1/laa/5A_COMB.pep:US-08-169-927-2
seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Wei
; APPLICANT: Dasch, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:

```

NAME: Spevack, A. David
 REGISTRATION NUMBER: 24,743
 REFERENCE/DOCKET NUMBER: 75,976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 295-6759
 TELEFAX: (301) 295-1022
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1612 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-169-927-2

alignment_scores:
 Quality: 243.50 Length: 1471
 Ratio: 0.362 Gaps: 72
 Percent Similarity: 45.683 Percent Identity: 19.511

alignment_block:
 US-09-303-518D-653 x US-08-169-927-2 ..

Align seg 1/1 to: US-08-169-927-2 from: 1 to: 1612

```

634 AATAACCGCGAAATTCATATCATATTCGAAGCGCATATCTTGCGCTCGT 683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 AsnSerArgSerSerSerSerSerSerSerSerSerSerSerSerSerSer 170
684 CGGTGGCATACCTTGGCACAATAAGATGATGAGTGGTGGCAGACGTAAC 733
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 .....AsnSerLysIleValAsnGly..GlyAsnGlyIleValAsnI 184
734 TA.....GGTAGCGAAAAATTAAACATAGCCCATATGCTTTTACCA 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
184 LeThrAsnGlyPheIleGlnValSerAspAsnThrPheIleGlyIleLys 200
778 ACAGACGGCTCATTTGGCGACAGTGGCTCCACCAATGTTATCATGATGCG 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
201 Thr.....IleAsnIleAspAspCysGlnGlyLeuMetPheAsnSerThrPr 216
828 CCNAAAGCAAAAGTGTTA.....A 847
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 AspAlaIleAsnThrLeuAsnLeuGlnIleGlyIleAsnThrIleAsnPr 233
848 TTAAATGGGTATTGCAACAGCAACCCCTATATAGGAAAAAGCAATGCG 897
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 heAsnGlyIleAspGlyThrGlyLysLeuValLeuValSerLysAsnGly 249
898 .....TTCCAGCTAGTCTCTAAAGATGAGTGTCTATGATGAAT 935
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
250 AlaAlaIleThrGlnPheAsnValIleGly..ThrLeuGlyGlyAsnLeuLysG 266
936 CTTTGTGAGATA...CCCATTCAGTATTCTAGCAACACATCAAAATG 982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 LylIleIleGlnLeuAsnThrAlaIleValIleGlyLysLeuIleSerLeu 282
983 GGA.....AATACCTTTTAAAGACACATATAATGGCGGAGGAAAA 1023
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
283 GlyGlyAlaIleAsnIleValIleGlyThr..AspAsnGlyIleGlyAsn 298
1024 ATCGATGCCAAACATAAACACATATTCCTACCTTATAGATTAAACACG 1073
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
299 AlaIleGly..... 301
1074 AACCGTTCATATGTTAAATGTTCTTATTCGAGACAGCAGAGAACCTG 1123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
302 .....PheIleValSerValAspAsnGlyAsnAlaIleAlaThrI 314
1124 TTTTATCATGCTGCAGTGGGGTCAACGTTATGACCCACAGATGAATAT 1173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
314 IeSerGlyGlnValIleValIleLysAsnMetValIleGlnIleSerAlaAsnAla 330

```

```

1174 GGAGAAATATTTTCCTTT.....ATTGACAAAGCAAAAGGT..... 1209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
331 GlyGlyGlnValIleThrPheGlnIleValAspValIleGlyGlyThr 347
1210 .....CAATTGATCTTACCACCAACATCA 1234
347 rThrAsnPheLysThrAlaSerSerLysValIleIleThrGlnIleValSer 364
1235 ACCAAGCGCG.....GGCGGTTTG..... 1254
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
364 snPheGlySerThrAsnPheGlyAsnLeuAspThrGlnIleValValPro 380
1255 .....TATTTGAGGGTAAATTTACGCTCTCGCTTAAAAACACGA 1295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 AspThrLysIleLeuLysGlyAsnPheIleGlyAspValLysAsnGln 397
1296 A.....ACGTGGCAGAGCGCGGCGCTTCATATGAGT 1326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
397 yAsnThrAlaGlyValIleThrPheAsnAlaAsnGlyAlaLeuValSer 414
1327 .....GATGGCAGTACCGTTACTTGCAAAAGTAAACGCG..... 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
414 IeSerThrAspProAsnIleAlaValIleThrAsnIleAsnAlaIleGlnAla 430
1360 .....GTGCGCAACACAGCG 1373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
431 GluGlyAlaGlyValValGlnLeuSerGlyIleHisIleAlaGlnLeuThr 447
1374 CCTGTCC.....AAATCGCGCAAGCGACGCTGCGTGG 1405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
447 gLeuGlyAsnGlyLysSerIlePheLysLeuAlaAspGlyThrValIle 464
1406 TTCACGCAAAAGGGAAC.....CAA 1428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
464 snGlyProValAsnGlnAlaLeuMetAsnAsnAlaLeuAlaIle 480
1429 GGCCTGCTCAGCGTGGCGACGCTAAAGTCACTTCAATACGACGCGCA 1478
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
481 GlySerIleGlnLeu...AspGlySerAlaIleIleThrGlyAspIleG 496
1479 CGATCAAGCGCAAAACACAGCTTAGGAATGGCTGGTGGCGCGCA 1528
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
496 yAsnGlyGlyValAsnAlaIleValGlnIleThrLeuAlaAsnAsp 513
1529 GGGGACGCGTG...CAACTGAATGCCGATATTCAGTCAACCCGAC... 1572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
513 IeSerLysIleLeuAlaLeuAspGlyAlaAsnIleIleGlyAlaAsnVal 529
1573 .....AAACTGTATTTGCGCTTCGCGGCGGACGTTGGATTGACOG 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
530 GlyGlyAlaIleHisPheGlnAlaAsnGlyLysThrIle..... 542
1617 GCATTGCGTTTCGTTCCACCCGATTCAAATATACGATGAGCGCGATGA 1666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
543 .....LysLeuThrAsnThrGlnAsnAsnIleVal 553
1667 TTGTCAACCAATCAAGACAAAGAAATCCACCGTTACCTTAGCGCAAT 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
553 al.....AsnPheAspLeuAspIleThrThrAspLysThrGlyVal 566
1717 AAAGATATTACTACCAACCGGCAATACCAACAACTGATGATAGCAAAAAAGA 1766
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
567 ValAspAlaSerSerLeuThrAsnAsnGlnThrLeu..... 578
1767 AATGCTTCAACAGGCTGTGGTGGAGAGAAAGATGCA...ACCAAAACGA 1813
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
579 .....ThrIleAsnGlySerIleGlyThrValValAlaAsnThrLysThr 594
1814 ACGGGCGGCTCAATCTGAATTCACCAACGGAAGAGCGGATCGCACTTA 1863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
594 euAlaGlnLeuAsnIleGly.....SerSerLysThrIle 605
1864 CTGCTTCCGGGCGAACAATTTAAACGCAATATACGCAAAACACGCG 1913

```

[illegible]

```

3387 GGAGGAGAGCAAAAAACGGGTGACGGCGATTAAGACACCGCTTGGCGA 3436
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1138 aglaspalaglinalasnalaThrThrThrThrThrThrThrThrThrThr 1155
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3437 AACAGCGGAGGAGGAAACCGG...CCGGTACACCGCTTCCCGCCG 3463
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1155 leglnaspasnalaasnalaasnpheserThrThrThrThrThrThr 1171
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3464 GCCCGCGCGCGCGCGGATTTGGCGCAACCGACCCCAACCGCAAC 3533
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1172 lleglnglyglyalaArg..... 1177
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3534 CCACCGGAGCGCGACCTGATACGCGTTATGCCAATAGCGGTTGAGTG 3583
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1178 .....PheAsnGlyThrLeuGlyAlaProA 1186
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3584 AATTTCGGCCGCTCAACAGCGTTTCGGCGTACAGGAGAAATTGGAC 3633
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1186 snPheAlaValThrThrThrThrThrThrThrThrThrThrThrThr 1200
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3634 CGCGTGTTCGCGAAGACCGCGCAACCGCTTTGGACAGCGGATCG 3683
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1201 ..... 1202
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3684 GGACACCAAAACACTACCGTTGCCAAGATTTCCGCGCTACCGCAACAA 3733
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1202 gaspser.....AsnGlnAspThr..... 1208
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3734 CGGACCTGGCCAAATCGGTATGCAGAAAACCTCGGACGCGGCGCGTC 3783
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1208 ..... 1208
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3784 GGCATCTGTTTTCGCACACCGGAGGAAACCTTCGACGAGCGCAT 3833
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1209 ...ValLeuThrArgThrAsnaspValLeuAsnValValThrThrAla 1224
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3834 CGGCAACTCGGCGCGCTTGGCCACGCTCCGTTTCGGGCAATACGCA 3883
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1224 lGlyAsnSerAlaIleAlaAsnAlaProIlyal..SerGlnAsnIleSer 1240
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3884 TCGGCA...GGTTCACATCGGATCGGCGGCGGCGGCTTTTGTATAC 3930
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1241 ArgCysLeuGlnSerThrThrAsnThrAlaAlaIleAlaIleAlaThr 1257
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3931 G.....GCACCTTTCAGACGCGCATCAGAGCAAAAT 3962
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1257 alysAspProSerAspValAlaThrPheValGlyAlaIleAlaIleAla 1274
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
3963 CCGCGCGCGCGCTGCTGCATACGCGCATTCAGCAAGATACCGCGCGGT 4012
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1274 hTrSerAlaAla.....ValThrThrValAsnLeuAsnAspThrGln 1288
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4013 TCGCGGATTCGGCATCGAACCGACATCGCGCAACGCGCTATTTCGTC 4062
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1289 ThrGlnaspLeuSerAsnArg..LeuGlyThrLeuArgThrLeuSer 1304
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4063 CAAAACGCGGATTCGATACGAAACGTCAATATGCGCACCGCGCGCT 4112
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1305 AsnAla.....GluThrSerAspValAla.....GlySe 1314
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4113 TGCATTCACCGCTACCGCGCGGCGCATTAAGCATTAATTCATTCAC 4162
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1314 rAlThrThrGlyAlaValSerSerGlyAspGlnAlaGlnValSerThr 1329
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4163 CGGCGCAACATTCATTCACGCGCTTATTTGACCTGCTCCTATACCAT 4212
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1330 .....GlyValThrPalalysProPheThrAsnIleAlaGlnGln 1343
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4213 GCCCGTTCCCGCG.....AAAGTCGACACGCGC..... 4239
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1344 LysIysGlyGlyIleAlaGlyThrLysAlaIleAlaThrThrThrThr 1360
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

4239 ..... 4239
1360 lGlyLeuAspThrLeuAlaSerAspAsnLeuMetIleGlyAlaAlaIleG 1377
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4240 .....GTCAATACCGCGCTATTCGGCGAGATTTCCGCAAAACCGCA 4284
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1377 lYlThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 1390
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4285 GCGGATGCGCGCTAAACCGCGCAATCAAGGTTTCACGCTGCGCTGCA 4334
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1391 .....GlyAspThrThrThrThrThrThrThrThrThrThrThrThr 1404
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
4335 CGCTGCC 4341
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
1404 rGlySer 1406
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: /cgn2/6/ptodata/1/aa/6B_COMB.pep:US-09-268-347-49

seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-49

alignment_scores:
Quality: 237.00 Length: 1232
Ratio: 0.422 Gaps: 59
Percent Similarity: 45.536 Percent Identity: 19.805

alignment_block:
US-09-303-518D-653 x US-09-268-347-49 ..

Align seg 1/1 to: US-09-268-347-49 from: 1 to: 2314

340 GCACATAACGGCGCTATACAAATGTGATTTGGTCGCGAGGAGCA 389
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
25 SerHisSerGlyGlySerSerSerThrAlaGlyGlnValGlySerSe 41
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
390 TCCCGATCAGCACCGCTTTCTTACCAAAATGTCAAAAGAAATTAATA 439
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
41 r.ProValIleArg..... 45
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
440 AAGCAGGAGCTAAGCGCATCCTATGCGGCGATTAATATATGCGCGGT 489
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
46 .....LeuThrArgValAlaThrLeuAlaIleLeuValIleGlyAl 59
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
490 TTGCACAAATTTGTCA...CAGATGCAGAACCTGTTGAGATGACAGTTA 536
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
59 arThrLeuAsnGlySerAlaThrAlaGlnAsnSerLysIleAlaPheG 76
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
537 TATGATGGGTGAAT.....ACGCTGANT 562
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
76 lYThrThrGlyAsnAsnAspAsnAlaSerAlaSerAsnGlnAlaSerIle 92
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
563 TAAATTAATTAATTCGATCGTTTCGATGCGAGCAGCAGCAATATTCG 612
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
93 AlalIleGlySerLeuAlaLysAlaHisAlaAsnGlnAlaIleAlaIle 109
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
613 CGGCTTGATGAGACGACCAACCAAT.....AACCGGAAAGTTTC 650
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :
109 yGly...SerLysProAspProArgAsnGlnAlaAlaAsnGlnLysAla 125
      |||: : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

651 ATATCATTTGTCAGCGCATATTCCTGGCTGCTGGCATACCTTTG 700
   : ||| :||| :||| :||| :||| :||| :||| :||| :|||
125 ySerHisAlaLysGlyLysGlnSerIleAlaIleGlyLysPValLeuA 142
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
701 CACAAATGGATAGGTGGTGGCAGCATCAACTTGTAGTAGGAAAAAT 750
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
142 IAGLnu.....GlyAspAlaSerIleAlaIleGlySerAspSpleu 155
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
751 .....AACATAGCCCATATGGTTTTCACCAAGAGGCTCAT 791
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
156 TyrLeuAspArgAsnSerThrAsnSerIleTyrProAsnGly..... 169
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
792 TGGCAGAGTGGCTCACCATGTATTATCTATGATGCCAAAGCAAAAGT 841
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 ..... 169
842 GGTTAATTAATGGGTATTTGCA.....ACAGGCAACCCCTATATAGA 885
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
170 ..LeuLeuSerThrLeuIleGlnAsnHisThrValLeuArgGlnIleArg 185
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
886 AAACGATGGCTCCAGCTAGTTCGTAAAGATTGGTTGATGATGAAT 935
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
186 AspSerAsnGlySerGlnLysTyrArgArgThr..... 196
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
936 CTTCCTGAGATACCCAT.....TCAGTATTCAGAACCCATC 976
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
197 ....AlaIleGlnGlyHisAlaSerThrAlaValGlyAlaMetAlaTyrA 212
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
977 AAAATGGAAATACCTTTTTCACGACAAATTAATAGCGCAGGAAAAATC 1026
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
212 IalysGlnHisPhe.....AlaAsnAlaPheGlyThrArg 223
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1027 GATGCCAAACATAACACTATTCCTACCTTAATAGATTAAAAACAGCAAC 1076
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
224 SerThrAlaGlnGlyAsnTyrSerGlnIleLeuAlaValGlyLeuThrAlaLysAl 240
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1077 CGTTCAATGTTTAATGTTCTTATCCGACAGCAGCAAGAAACCGTTT 1126
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
240 agLysGlyTyrThrIleAlaIleGlySerAsnAlaGlnAlaIleAsnT 257
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1127 ATCATGCTGCA.....GGT 1140
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
257 TyrGlyAlaLeuAlaLeuGlyAlaAspThrArgValAspLeuAspTyrGly 273
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1141 GGGGTCAACAGTTATCGACCCACAGC.....AATAATGAGAA 1178
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
274 IleAlaLeuGlyTyrGlySerGlnIleLeuAsnAsnAsnAsnAsnAsn 290
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1179 AAATATTTCTTATTTGACAAAGAAAGTAATTGATTAATTAACAGGTC 1228
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
290 nAsnLysAlaTyrValProGlnGlyAsnGlySerAsnIleLysSerSerL 307
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1229 ACATCAACCAAGCGCGCGGCTTTGTATTTGAGGGTAATTTACGGTC 1278
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
307 ysaLathr.....GlyAsnGlyLeuPheSerIleGlySerSerThrIle 321
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1279 TCGCCTAAATAACAACGAAACGTGGCAAGCGCGGCTTCATATCAGAGA 1328
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
322 LysArgLysIleIleAsnVal.....GlyAlaGlyTyrGlnAspThrAs 336
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1329 TGGCAGTACCGTTACTTGG.....AAAGTAAACGGCTGGCAAAAGCAGC 1372
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
336 pAlaValAsnValAlaGlnIleuLysAlaValGlnAsnLeuAlaLysArg 353
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1373 GCCTGCCAAATTCGGC.....AAAGCAGCGCTGCTGTCAAAGCCAAA 1416
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
353 IiIleThrPheLysGlyAspAspAsnGlyThrGlyValLysLysLeu 369
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1417 GGGGAAACCAAGCGCTCGTCAGCGTGGCGAGTAAGTCATCTTACA 1466
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
370 GlyGlu.....ThrLeuThrIleLysGlyLysGlnGlnAlaAs 383

```

```

1467 TCAGCAGGCGGACGATCAAGCAAAAAACAGCCCTTAGTAATGGCT 1516
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
383 pLysLeuThrAspAsn.....AsnAsnIleGly 393
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1517 TGGTACGCGGACGGGAGC.....GTGCACTG..... 1545
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
393 alValThrAspAsnAsnThrIleLysValLysLeuAlaLysAsnLeu 409
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1546 .....AATGCCATATTCAGTTCAACCCGACAACT 1577
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
410 SerGlyLeuGlnThrValSerThrLysAsnLeuThrAlaSerGlnLysVal 426
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1578 CTATTCGGCTTTCGCGGACGCTTGGATTGTAAGCGCATTCGTTT 1627
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
426 lThrValGlySerGlyAsnAsnThrAlaGlnLeuGlnSerGlyLeuT 443
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1628 CGTTC.....CACGCGATTCAAATAC 1650
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
443 hrPheThrProThrThrAsnAlaSerThrAspLysThrValTyrGlyThr 459
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1651 GATGAAGGCGGATGATTCACACCAATCAACAGCAAAAGAACCCGCT 1700
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
460 AspGlyLeuLysPheThrAspAsnSerAsnThrAlaLeuGlnAspThrTh 476
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1701 TACCATTTACA.....GGCAATTAAGATTAATTAACA 1732
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
476 ArgIleThrLysAspLysIleGlyPheSerAsnLysAlaGlyThrValA 493
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1733 CCGGCATTAACACACTTGGATAGCAAAAAAGAAATTCCTTACACGGT 1782
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
493 spGlnAsnLysProTyrLeuAspLysAspLysLeuLysValGlnLysn... 508
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1783 TGGTTTGGCGAAGAAAGATGCACCAACCAAGACGGCGGCTCAATCTGA 1832
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
509 .....SerThrLeuAsnAsnGlyGlyLeuThrValAs 519
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1833 TTAACCAACGGAAGAACGCGATCGACTTACTGCTTCGCGGAGAA 1882
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
519 n.....AsnThrIle.....GlyGlySerA 526
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1883 AT.....TTAAACGGCAAT 1896
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
526 snLysGlnIleGlnValGlyAlaAspGlyIleLysPheAlaAspValAsn 542
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1897 ATCAGCAACCAACACGCAAACTGTTTTCACGCGGACGACGACCGCA 1946
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
543 ValAsnValSerAsnAlaIleLysPheGlyThrThrArgIleThrGlnG 559
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1947 CGCTTCAATTCATTTAGAAAGCGGCTGTCAAAAATGGAAGTATCCAC 1996
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
559 uGlnIleGlyPheAlaAspAla.....A 567
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
1997 AAGGAGAATTCGTGGGACACGATTTGATTCGACCGCATTTAAAGCG 2046
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
567 spGlyLysValAspLysLysSerProTyrLeuAspLysLys..... 580
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
2047 GAAACCTTCATATTCAGGCGGACAAAGCGGTGTTCCCAATGTTCG 2096
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
581 .....GlnLeuGlnValGlyValLysIleThrLysAspSerGly 594
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
2097 CAAAGTGAAGCGCATTTGACATTAACGATTCACGCCAACGACTT... 2142
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
594 ylleAsnAlaGlyAspGlnLysIleSerAsnValLysAspAlaThrAsp 611
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
2142 ..... 2142
611 spThrAspAlaValThrTyrLysGlnLeuLysGlnValGlnIleAspAla 627
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
2143 .....TTGCGTGTGCAACCGCAT.....CAAAGCCA 2168
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
628 AspGlyAlaLeuGlnSerPheSerIleArgAspGlnLysGlyGlnLys 644
   :||| :||| :||| :||| :||| :||| :||| :||| :|||
2169 CACAATCTGTACAGTTTCGACTGACGCGGTCTGACAAAGTTTACCGAAA 2218

```



```

644  |||||  ::  ::  ::  |||  |||  ::
ethrllserasnleutyser...asnglyasnthrproasnthrheg 660
2219  AACCATTACC.....GACGATAAGGATGCTTCATTGAGCAAG 2259
:::|||||  ::  ::  ::  |||  ::
660  luthrllrthrphaelaglyglusnnglyle.....serlleserasn 674
2260  ACCGACATCAGAGCAATGTCAGCTTGCCGATCAGCTCATTTAAATCT 2309
:::|||||  ::  ::  ::  |||  ::
675  Aspllealalysglylvalysvalysval..... 683
2310  CACAGGATTCACACATCAACGGC.....AATCTAGTGAG 2347
|||||  ::  ::  ::  |||  ::
684  ...glylleasprroilleasnnglyleuthrThrProlysleuthrvalg 699
2348  GCGGAGAC.....ACGCACTATPACGTTACCGCGCAACGCCACC 2385
|||||  ::  ::  ::  |||  ::
699  lysrAspLysaspLyspLysThrGlnleuValIleGlnIvalAlaser 715
2386  CAAACGGCAACCTGACCTGTGGCAATGCCAGACACATTTAAT.. 2433
|||||  ::  ::  ::  |||  ::
716  gLysannglyThrLysasnIleleArnglyValSerProthrLeuproSe 732
2434  .....CAAGCCATTAACGCAACGCAACATCGG 2461
732  rllrthrsnalaglyglvalargThrThrGlnIleGlnlysnthrleAr 749
2462  CTTGCGAC.....AAT 2472
:::|||||  ::  ::  ::  |||  ::
749  hserAspLysaspLysserLysAlaAlaserIleGlyAspIleleuasn 765
2473  GCTTCATTATCTAAGCAACACGCCGTCACAAACGGCAGCTGACG.. 2520
:::|||||  ::  ::  ::  |||  ::
766  ThrGlyPheasnleuLysasnSerasnSerValIleGlyPheValSerThr 782
2521  .....CTTTCGACACAGCT...AAGCCAAACGTAA 2548
:::|||||  ::  ::  ::  |||  ::
782  rlyrasnThrValAspPheIleaspLysnAlaThrThrAlaLysValr 799
2549  GCCATTCCGCACTCAACGGC.....AATGTCTCCCTA 2580
:::|||||  ::  ::  ::  |||  ::
799  hTyrAspGlnThrAsnGlnThrSerLysValThrTyrAspValAsnVal 815
2581  GCCGATTAAGCACTATTCATTGTAACAGCCGCTTACCGGAATAAT 2630
:::|||||  ::  ::  ::  |||  ::
816  AspGlnLysThrIleGlnleuThrThrGlyAspAsnGlyLysThrAsnLys 832
2631  C.....ACGGCGCGCAAGATA 2647
:::|||||  ::  ::  ::  |||  ::
832  eGlyValLysThrThrThrleuthrThrThrAsnAlaAsnGlyLysAla 849
2648  CCGCATTACCTTAAGACAGCAATGGACGCTGCCGTCGGCAGCGAA 2697
:::|||||  ::  ::  ::  |||  ::
849  hrasnpheSerThrThrAspAsnAspAlaLeuValAsnAlaLysAspIle 865
2698  TTAGGCAATTTAAAC..... 2712
|||||  ::  ::  ::  |||  ::
866  AlagLusnleuasnThrleuAlaLysGlnIleHisThrThrLysGlyThr 882
2713  .....CTTG 2716
882  rAlaAspThrAlaLeuGlnThrPheLysValLysLysAspGlyAlaThrA 899
2717  ACAACGCCACCATTAACATCAATCCGCTATGACACGATGGCGCGCG 2766
:::|||||  ::  ::  ::  |||  ::
899  spAspGlnThrThrValGlyLys.....AspGly 909
2767  GCGCAACCGCGCAGTGGCGAGATGCGCGCGCGCGCTTCGCGCCCTTC 2816
:::|||||  ::  ::  ::  |||  ::
910  ThrGlnsnnglyLysThrVal..... 916
2817  CCTATTATCCGTTAGCGCGCACTTGCGAGAAATCCCGTTTCAACGCG 2866
|||||  ::  ::  ::  |||  ::
917  .....AsnThrL 919
2867  TGACGTTAAACGGCAATTAAGCGGTCAGGAGAAATTCGCTTTATGTCG 2916
:::|||||  ::  ::  ::  |||  ::
919  eutLysleuLysGlyGlu...AsnGly..... 926
2917  GAACCTTCGGCTACCGCAGCGGCAATTGAAGCTGGCGGAAGTTCCGA 2966
:::|||||  ::  ::  ::  |||  ::
927  .....LeuthrValAlaThrAsnLysAs 934
2967  AGCCTACTACCTTGCGCTGTACACATATACCGCAACGACCGATGACG 3016
|||||  ::  ::  ::  |||  ::
934  pGlyThrValThrPheGlyLleasnThrGlnSerGly..... 946
3017  TCGAGCAATTGACGATGAGGAGAAAGACACACACCGCTGTCCGAA 3066
:::|||||  ::  ::  ::  |||  ::
947  .....LeuLysAlaGlyAspSerThrThrLeuAsnLys 957
3067  .....AATCTTAATTTACCTGCAACAAACGACACGTCGATCCGG 3107
:::|||||  ::  ::  ::  |||  ::
958  AspGlyLeuSerIleLysasnProAlaserAsnGlnIleGlnValG 974
3108  CGCA.....TGGCGTTATCAGCTTATCCGCANA..... 3135
|||||  ::  ::  ::  |||  ::
974  yAlaAspGlyValLysPheAlaLysValAspLysGlyAsnSerSerThr 991
3136  .....GACGGGAGTTCGCTGCAT..... 3156
|||||  ::  ::  ::  |||  ::
991  LylleAspGlyThrSerArgIleThrLysAspGlnIleGlyPheThrGly 1007
3157  .....AATCGGTCACAAACACAGCTTTCGCAACACT 3191
:::|||||  ::  ::  ::  |||  ::
1008  AlaasnGlySerLeuAspThrThrLysProHisLeuThrLysAspLys 1024
3192  CCGCAAGCGCGAGAAACAGACGCCGCTTGACGCGCAAAACAGCA... 3237
|||||  ::  ::  ::  |||  ::
1024  u...LysValIleGlnValGlnIleThrAsnThrGlyLleAsnAlaGly 1040
3238  .....CAACTGCGCGCCAAACACAGCGCGGAAAAACACACGCGCAAGC 3282
:::|||||  ::  ::  ::  |||  ::
1040  LylLysLysIleThrAsnIleGlnSerGly.....AspIleThrGlnAsn 1054
3283  CTTGACGCGCTGATTCGCGCGCGCGC.....AATGC 3314
:::|||||  ::  ::  ::  |||  ::
1055  SerAsnAspAlaValThrGlyGlyAlaValAlaThrAspLeuLysThrGln 1071
3315  CACCGAAGCGAGCAAAAGTGTGCGGACCGCGCGCGAGCA 3357
:::|||||  ::  ::  ::  |||  ::
1071  uGlnSerLysIleAsnSerAlaAlaLysThrAlaGlnAsnSer 1085
seq_name: /cgn2_6/ptodata/1/aa/5b_comb.pep:US-08-968-685A-10
seq_documentation_block:
: Sequence 10, Application US/08968685A
: Patent No. 6214981
: GENERAL INFORMATION:
: APPLICANT: TUCKER, KENNETH
: TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
: TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ. ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

```

```

alignment_scores:
Quality: 228.50      Length: 1754
Ratio: 0.290         Gaps: 88
Percent Similarity: 44.983      Percent Identity: 19.327

```

alignment_block:

US-09-303-518D-653 x US-08-968-685A-10 ..

Align seg 1/1 to: US-08-968-685A-10 from: 1 to: 2123

```

122 GGGGGGACACACTTATTCGGCATCACTACCAATATCTGCGACTT 171
    ||| ||| : : : : : : : : : : : : : : : : : :
590 GYAlAAspIyIleThrPheThrAspIleSerAsnSer..... 602
172 GCGGAATAAAGCAAGTTTGGCGGGGGGCAAAAGATATGAG... 217
    ||| ||| : : : : : : : : : : : : : : : : : :
603 ...LysProGlyAlaGlyIleIleLysnThrThraGlyIleThraGlyAspG 618
    ||| ||| : : : : : : : : : : : : : : : : : :
218 .....TTTACACAAAAAAGGAGAGTTGGTCGGCAATGATGAGCA 259
    ||| ||| : : : : : : : : : : : : : : : : : :
618 LylIleGlyPheAlaAsnThrGlySerLeuAspAlaAsn..... 631
260 AACCCCGATGATGATTTCTGTGCTATCCCGTAACGCGGTGCGCA 309
    ||| ||| : : : : : : : : : : : : : : : : : :
632 LysProArg..LeuThrProThrGlyIleAsnAlaGly..LysGlu 646
310 TTGGCGGGGCAATATATTTGTGACGCGGCACATTAACGGGGCTATAA 359
    ||| ||| : : : : : : : : : : : : : : : : : :
647 LeuThrAsnValGlnSerAlaIleAsnProAlaThrAsnGlyGly... 661
360 CAATGTGTTTGGTGGCGAGGAGCAATCCGATCAGCAGCGCTTTT 409
    ||| ||| : : : : : : : : : : : : : : : : : :
662 ..GlnLeuAspPheMetAsnArgLeuSerThrAlaAsnThrGluLysSerG 678
410 CTTACCAATTTGTGAAGAATAATATTAAGAAGGAGGACTAAGGCCAT 459
    ||| ||| : : : : : : : : : : : : : : : : : :
678 LysSerAlaIleThrIleLysAspLeuArgLeuSerGlnValProLeu 694
460 CCTTATGGCGGATATCATATGCGCGCTTGGCAACAATTTGACACAGA 509
    ||| ||| : : : : : : : : : : : : : : : : : :
695 ThrPheAlaGlyAspThrGly...ProAsnValThrLys..... 706
510 TGCAGAACCTGTGAGATGACCACTTATATGATGGGTGAATAATACGCTG 559
706 .....
560 ATTTAATAAATACCTGATGCTGTGCAATGAGAGGAGCAACAATAT 609
    ||| ||| : : : : : : : : : : : : : : : : : :
707 .....LysLeuGlyIleIleLysValLysGlyGlyLysThr... 719
610 TGGCGGTGTGATGAAGCAAGCAACCAATATAC..... 639

```

```

720 ..ThraIAspAspLeuThrLysAsnAsnIleGlyValAlaIAspSe 735
    : : : : : : : : : : : : : : : : : :
640 .....CGCGAAAGTTCAATATTCGACAGGCGCATATCTGGCTCG 682
725 rThraSpAsnSerLeuThrValLysIleAlaLysThrLeuSerAspLeuA 752
663 TCGGTGGCAATATCTTTGACAAAAATGATCAGGTGGTGGCAGCTCAAC 732
    : : : : : : : : : : : : : : : : : :
752 spAlaValAsnThrLysThrLeuThrAlaSerAspLysValThrValAsp 768
733 TTAGGTAGCGAA.....AAATTAACATAGCCCATATGCTTTTACC 776
    ||| ||| : : : : : : : : : : : : : : : : : :
769 SerGlyAsnAsnThrAlaLysLeuGlnAsnGlyIleAlaLeuAspGlyThr 785
777 AACAGAGGCTCATTTTGGCAGAGTGGCTCCACCAATGTTATCTATGATG 826
    : : : : : : : : : : : : : : : : : :
785 sGlnAsnThrGlyAlaThrProAlaThrAsnSerLysThrIleThrGlyY 802
802 CCCAAAGCAAAAGTGTGTA..ATTATGGGTATTTGCAACAGCGCAC 873
    : : : : : : : : : : : : : : : : : :
827 aLAspGlyLeuLysPheThrAspAsnAsnGlyIleAlaLeuAspGlyThr 818
    : : : : : : : : : : : : : : : : : :
874 CCTATATAGCAAAAGCAAT...GGCTCCAGCTGATGTCGTAAGATTG 920
    ||| ||| : : : : : : : : : : : : : : : : : :
819 ThrThrIleThrLysAspLysValGlyPhe.....AlaLysGlnAsp.G 833
921 GTTCTATGATGAATCT.....TTGCTGGAG 946
    ||| ||| : : : : : : : : : : : : : : : : : :
833 LysLeuAspLysSerLysProThrLeuAspLysAspLysLeuLysVal 849
947 ATACCATTCAGTATCTACGAACACATCAAAATGGAATATCTTTT 996
850 GYGlValAlaGluIleThrThrAsnGlyIleAsnAlaGlyGlyLysAlaI 866
997 AACG.....ACATATATATGCGCGAGGAGCAAAAGATGCCAA 1034
    ||| ||| : : : : : : : : : : : : : : : : : :
866 ethrGlyLeuSerAsnThrLeuThrAspAlaThrAsnAlaThrThrGlyH 883
1035 ACATTAACACTATCTCTACTTATAGATTA.....AAACAGCAAC 1077
    ||| ||| : : : : : : : : : : : : : : : : : :
883 IAspVal...ThrGlnLeuGlyIleValAspSerThrAspLysThrArgAla 898
1078 GTTCAATGTTTAAATGTTCTTATCCGACAGACAGACAGAACCACTGTTA 1127
    ||| ||| : : : : : : : : : : : : : : : : : :
899 AlaSerIleGlyAspVal..... 904
1128 TCATGCTGAGGTGGGCTACACAGATTATCGACAGACTGAATATGAG 1177
    ||| ||| : : : : : : : : : : : : : : : : : :
905 .....LeuAsnAlaGlyP 909
1178 AAAATATTTCTTTATGACAAAGAAAGTGAATGATATCTTACACAG 1227
    ||| ||| : : : : : : : : : : : : : : : : : :
909 hAsnLeuLysAsnAsnGlyAspAlaLysAspPheValSerThrTyAsp 925
1228 AACATCAACCAAGCGCGCGGCTTTGTATTTAGAGTATTTTACGT 1277
    ||| ||| : : : : : : : : : : : : : : : : : :
926 ThrValAsp.....PheIleAsnGlyAsnAlaThrThr 936
1278 CTCGCTTAAAAACAAGCAACGTCGCAAGCGCGGCTTCATATAGT. 1326
    ||| ||| : : : : : : : : : : : : : : : : : :
936 rAla.....LysValThrTyAspGlyLysAlaSerLysValAlaIat 950
1327 .....GATGGCAGTACCGCTTACTTGAAGAAGTAACGCGCTG 1362
    ||| ||| : : : : : : : : : : : : : : : : : :
950 rAspValAlaValAlaSpIyThrThrIle.....HisLeuThrIly... 963
1363 GCAAAAGCCCGCTGTCCAAATGCGC...AAAGCAGCGCTGTGATCA 1409
    ||| ||| : : : : : : : : : : : : : : : : : :
964 AlaAspGlyAsnLysAsnGlnIleGlyValLysThrThrThrLeuThrIly 980
1410 A.....GCCAAAGGGGAAC.....CAAGCTCGGTGACGCTGGCGG 1447
    : : : : : : : : : : : : : : : : : :

```

```
980 sThAspAlaLysGlyAspLysAlaIleAsnPhSeValAsnSerGlyA 997
1448 ACGGTAAGTCATCTTAGATCAGCGCGACATCAAGCGMAAAACAA 1497
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 spAspLysAlaLeuIleAsn.....AlaLysAsp 1006
1498 GCCTTTACGTAATCGCTTGTCAGCGC.....AGGGG 1552
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1007 lLeAlaAspAsnLeuAsnThrLeuAlaGlyIleArgAsnThrLysGI 1023
1533 GACG.....GTGCACTGA 1546
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 YThrAlaSpThrAlaLeuGlnThrPheGlnValLysLysValLysGI 1040
1547 ATGCGGTAATCACTTCAACCCGCAAACTCTATTGCTTGCGCGC 1596
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1040 snGlyAspAspAsnAspAlaAspThrIleThrValGlyLysAspAla 1056
1597 GGACGTTTGATTTGACGGCGCATTCGTTCCACCGCATTCAMAA 1646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1057 LysThrAsnGlnValAsnThrLeuLysLeu..... 1066
1647 TACCGATGAGGGCGCATGATGTCACCAATCAAGCAAAAGATCCA 1696
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1067 .....LysGlyLysAsnGlyLeuAspIleGlnThrAsnLysAspGI 1081
1697 CCGTTACCAT.....ACAGGCATTA 1719
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1081 hValThrPheGlyIleAsnThrGlnSerGlyLeuLysAlaGlyAsn... 1096
1720 GATATTACTACACCGGCAATACAACTGATAGCAAAAAGAAAT 1769
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1097 .....AsnThrThrLeuAsnAsnGlyLeu...SerIleLysAsnThr 1110
1770 TCGCTACAAAC.....GGTTGTTGCGCAGAAAGATGCCAA 1804
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 rAlaGlyAsnGlnGlnIleGlnValGlyAlaAspGlyValLysPheAlaL 1127
1805 CCAAAACGACGGCGGCTCATCTGATTTACCAACCGGAAGACGGAT 1854
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1127 ysValAsnAsnGlyValVal..... 1133
1855 CGCATTTTACTGCTTCCGCGCAACAATTTAAAGCAATATCAGCA 1904
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 .....GlyAlaGlyIleAspGlyThrThrArgLI 1143
1905 AACAAACGGCAACTGTTTTCAGCGC.....AGAC 1936
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1143 eThrArgAspGlnIleGlyPheAlaGlyThrAsnGlySerLeuAspLys 1160
1937 CGACACCGCAC...GCCTACATCATTTAGAGCGGGGTGCAAAATG 1983
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1160 eLysProIleLeuSerLysAspGlyIleAsnAlaGlyLysLysIle 1176
1984 GAAGGTATCCACAGAGAAATCTGTGGGACAGACGATTGATGCAGCG 2033
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1177 ThrAsnIleGlnSerGlyIleAlaGlnAsnSer..... 1188
2034 CACATTTAAAGCGAAACTTCATATTCAGCGCGACAGCGGTGTT 2083
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1189 .....AsnAspAlaValThrGlyLysIleTyr... 1198
2084 CCGGCATATGTTCCAAAGTGAAGGCGATGGCATTTAAACATCAAGCG 2133
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1199 .....AspLeuLysThrGlnLeuGlnLysLysIleSerSerThrAla 1212
2134 CAAGCAGCT.....TTGCTGTGCGCAGCGCATCAAG 2165
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1213 LysThrAlaGlnAsnSerLeuHisGlnPheSerValAlaAspGlnGlnI 1229
2166 CCACACATCTGTACACGTTGCGGACTGGAGGGGCTG...ACAAGTGTGA 2212
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 ysAsnAsnPheThrValSerAsnProTyrSerSerTyrAspThrSerLysT 1246
2213 CCGAAAAACCATTTACC.....GACGATTAAGTGAATTCCTTCATG 2253
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1246 hSerAspValIleThrPheAlaGlyGlnAsnGlyIleThrThrLysVal 1262
2254 AGCAAGACCGACATCAGA..... 2271
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1263 AsnLysGlyValValAlaArgValGlyIleAspGlnThrLysGlyLeuThr 1279
2272 .....GGCATGTCAGCTTG 2287
1279 rProLysLeuThrValGlyLysAsnAsnGlyLysGlyIleValIleAsp 1296
2288 CCGATCGCTCATTTAAATCTCAGAGACTTCCACACTCAAGCGCAAT 2337
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1296 eArgLysnGlyLysAsnThrIleThrGlyLeuSerAsnThrLeuAlaAsn 1312
2338 CTTAGTCAGCGGAGACAGCCACTATACGTTACCGCGCACCGCCACA 2387
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1313 Val.....ThrAsnAspLysGlySerValArgThrArgI 1324
2388 AACGGCAACCTC.....AGCTCG 2407
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1324 uGlnGlyLysIleIleLysAspIleAsnLysThrArgAlaIleSerIleV 1341
2408 TGGGCAATGCCCAAGCAACATTTAATCAGCCACATTAAGCGCAACCA 2457
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1341 AlaSpValLeuSerAlaGlyPheAsn.....LeuGlnGlyAsnGly 1354
2458 TCGGCTTCGACAACTGCTCA.....TTTAATCTAGCAACAA 2495
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1355 GluAlaValAspPheValSerThrTyrAspThrValAsnPheAlaAspGI 1371
2496 CGCGGTACAAACGCGCAGTGTGACGCTTCCGACACGCTTAAGCGCAAC 2544
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1371 ysAlaIleThrThrAlaLysValThrTyrAspAspThrSerLysThrSerL 1388
2544 ..... 2544
1388 ysValValTyrAspValAsnValAspAspThrThrIleGluValLysAsp 1404
2544 ..... 2544
1405 LysLysLeuGlyValLysThrThrThrLeuThrSerThrArgLI 1421
2544 ..... 2544
1421 aAsnLysPheAlaLeuSerAsnGlnAlaThrGlyAspAlaLeuValLysA 1438
2545 .....GTAAAGCAT...TCCGCACTCAACGCAATGTCCTCA 2580
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1438 lAserAspIleValAlaHisLeuAsnThrLeuSerGlyAspIleGlnThr 1454
2581 GCC..... 2583
1455 AlAlysGlyAlaSerGlnAlaAsnSerSerAlaGlyTyrValaAspAlaAs 1471
2584 ....GATAAGCAGTATTC...CATTTGAAACAGCGCTTTACCGGAA 2626
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1471 pGlyAsnLysValIleTyrAspSerThrAspAsnLysTyrGlnAlaL 1488
2627 AATCAGCGCGCGCAAGATACGGCATTCACCTTAAGAGACGCAATGG 2676
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1488 ysAsnAspGlyThrValAspLysThrLysGluValAlaLysAspLysLeu 1504
2677 ACGGTGCGGTGCGGACGCAATTAGCAATTTAACTTTGACAAAGCCAC 2726
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1505 ValAlaGlnAlaGlnThrProAspGlyThrLeuAlaGlnMetAsnValLy 1521
2727 CATTAACCTCAATTCGCTATGACAGCATGCGGAGCGCGCA... 2772
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1521 sSerValIleAsnLysGlnGlnValAlaAsnAspAlaAsnLysLysGlnGlyI 1538
```

```

2773 .....ACGGCAGTGGCGGAGATGGCGCGC 2799
1538 leasngluaspasnalaphevalysglyleuglulysalalaserasp 1554
2800 CGCGTGGCGCGCTTCCCTATTATCCGT..... 2829
1555 Asnlysthrlyspasnalalavalthrvalglyaspheuasnalavalal 1571
2830 ....ACGGCG.....CCAACTTGGCAGAAATGCC 2854
1571 aglnthrproleuthrphelaglyaspthrlythrthrvalylslysl 1588
2855 GTTCAACACGCTGACGGTAAACGGCAATTTGAACGGTCAG...GAACA 2901
1588 euglyglthrleuthrlelysgly.....glylnthraspthr 1601
2902 TTCCGCTTATGCGAACTCTTGGCGTACGCGCGC..... 2940
1602 Asnlysthrthrspasnasnileglyalavalaglythraspglyph 1618
2941 .AAATTGAAGCTGGCGAAAGTTCGGAAGCATTACACCTTGGCTGTCA 2989
1618 eThrVallysthrleualalysaspthrleuthrasnleuasnserval..... 1632
2990 ACAATTACCGGACGAAACCGGTAACTCGAGCAATTGACGTAAGTG... 3036
1633 .Asnalaglyglythrlysthrleaspaspysglyvalserphevalasp 1648
3037 .....GAAGAAAAGACACACACCGCTGCCCAAAATCTTAATTT 3077
1649 Sersesterglylnalalysalalaspthrproval..... 1659
3078 CACCTGTCAAAACGACACGTCGATCGCGCGCATGCGTTATCAGTTA 3127
1660 ....LeuSerAlaasnlyleuasnspleuugly.....lysval 1672
3128 TTCGCAAAAGACGGGAG...TTCGCGCTGCATTAATCCGTCAAAAGACAA 3174
1672 leSerasnvalglylysglythrlyspaspthraspalalaspvalgln 1688
3175 GAGCTTCCGAC.....AAACTGGCAAGCGCGGAGCAAAAC 3209
1689 Glnleuasnlyvalargasnleuleuglyleuglyasnalelysnas 1705
3210 AGAGCGCGCTTGAACGCAAAACGACACACTTGGCGCAACACAGC 3259
1705 pasnalal.....Aspelyasnlnvalasnilealaspillelylysa 1720
3260 CGAATAAAGACACGCGCAAAAGCTTGACGCGCTGATGGCGC..... 3303
1720 spProasnsrlySerSerSerasnargthrvalillelysalaglythr 1736
3304 .....GGCGCAATGCCACGCAAAAGCGCAAAAGTTCG 3338
1737 valleuglyglylyslnasnaspthrlylulysleualatthr.... 1751
3339 CGAACCGCGCGCGCAGCGGCGGAAATGCCGCGATTATGACGCGC 3388
1752 .....glyglyvalglnvalglyval..... 1758
3389 AGAAGAGAAAAACGGGTGCAAGCGGATAAAGACACGCTTGGCGAAA 3438
1759 .....Asplysasp..... 1761
3439 CAGCGGAGAGGAAACCGCGCGGTACACACGCTTCCCGCGCGCG 3488
1762 .....gl 1762
3489 CGCGCGCGCGGATTTGCCCAACCGAGCGCCACGCAACCCGAA... 3537
1762 yasnalaasnlyaspleserAsnvaltrpvalylsthrlnlyspaspg 1779
3538 ..CGCGAGCGCACTGATCAGCGCTTATGCCAAATAGCGGTTGAGTAA 3585

```

```

1779 lySerlysthrlysalaleuleualatthrtyrasnalalaglycIntprasn 1795
3586 TTTTCCGCGCAGCTCAACAGCGTTTTCGCGCTACAGACGAAATTTGACCG 3635
1796 Tyr...LeuthrasnspnProalaglualelle.....Aspar 1807
3636 CGTGTTCGCGAA..... 3648
1807 glleasnglulnglylleargppherhlsvalasnaspglyasnclng 1824
3649 .....GACCGCGCAACGCGTGTGGACAAAGCGGCATCCGGAC 3687
1824 lndProvalvalglnglyargasnlylleaspserseral.....Ser 1838
3688 ACCAAACTACCTACCGTTCGCAAGATTTCCGCGCTACCGGCAACAAACGA 3737
1839 Glylyshisservalalalileglypheglnalalysalalaspglylual 1855
3738 CTTGCGCCAAATCGGTATGCAAAAAACCTCGGACGCGCGCTGGCA 3787
1855 alavalalalleglyarglnthrlnaglyasnnglnsertillealai 1872
3788 TCCTGTTTCGCAACACCGACCGGAAC...ACCTTGACGACGCGCATC 3834
1872 leglyaspasnalaglnalatrnglyaspnlserillealai 1888
3835 GGCACCTCGGACAGGCTTGCACGCGTCCGCTTTCGCG..... 3873
1889 Glyasnvalvalalaglyshisserglyalalleglyasproserth 1905
3874 .....CAATACGCGATCGGACAG..... 3891
1905 rVallysalaspasnseryserValglyAsnasnasnlnphethra 1922
3892 .....TTCGACATCGGC..... 3903
1922 spAlatThrlnthrAspvalPheglyvalglyasnlnlethrvalthr 1938
3904 .....ATCAGCGCGCGCGGCTTTAGTACGCGGACGCTTC 3941
1939 GluserAsnservalalaleugly.....Seranseralaliese 1952
3942 AGACGCGATCAGAGGCAAAATCCGCGCGCTCGCATTAACGCGCATTC 3991
1952 rAlalely.....Thrlsalaglythrg 1960
3992 AGGCAAGATACCGCGCAGTTTCGGC..... 4017
1960 lndAlalyslySerAspelythrAlaglythrthrthrAlaglyAla 1976
4018 .....GGATTGCGCATGCAACCGCATCGCGGACGCGC 4052
1977 ThrlythrVallysglyphealaglyglnthrvalalaglyalavalse 1993
4053 CTATTTCGTCCAAAAAGCGGATTACCGATACGAAACGCTAATATC.... 4098
1993 rValglyAlaserglyAlagluargrgileglnasnvalalalaglyg 2010
4098 ..... 4098
2010 lvalalserlatthrSerThrAspAlavalasnlyserGlnleuYrlyys 2026
4099 GCCACCGCGCGCTTGA..... 4116
2027 AlatrhrnglyllealaasnAlatThrasnGlnleuasphlsarglleh 2043
4117 .....TTACCGCTACCGCGCGGCGCATTAAGCAGATTAATTCATTCA 4159
2043 sglAsnngluasnlysalasnalaglylleaser..... 2055
4160 AACCGGCAACAAATTTCCATCAGCGCTTATTGAGCCGTCTTATACC 4209

```



```

2830 .....ACGCG.....CCACTCG 2844
      |||||
1493 snAlaValAlaGlnThrProLeuThrPheAlaGlyAspThrGlyThrThr 1509
      |||||
2845 GCAGAAATCCGCTTCAACACGCTGACGGTAACGCCAATTCAGCGTCA 2894
      |||||
1510 AlAlaGlyLeuGlyGlnThrLeuThrIleGlyGly.....GlyG 1523
      |||||
2895 G...GGAACATTCGCTTATGTCGAACTTCGCTACCGACGCGC 2940
      |||||
1533 nThrAspThrAsnLysLeuThrAspAsnIleGlyValAlaIleGly 1540
      |||||
2941 .....AAATGAACTGGCGGAAAGTTCCGAAGGCACTTACAC 2979
      |||||
1540 hrAspGlyPheThrValLysLeuAlaLysAspLeuThrAsnLeuAsnSer 1556
      |||||
2980 TTGGCTGTCACATACCGGCAACGCCCTAGTCTCGAGCAATGAC 3029
      |||||
1557 Val.....AsnAlaGlyGlyThrLysIleAspAspLysGlyAlaSe 1570
      |||||
3030 GGTAGT.....GAAGAAAAGACACACACCGCTGTCGAAA 3067
      |||||
1570 rPheValAspSerSerGlyGlnAlaLysAlaAsnThrProVal..... 1584
      |||||
3068 ATCTTAATTTACCTCGAANAACGACGTCGATGCCGCGCATGGCGT 3117
      |||||
1585 .....LeuSerAlaAsnGlyLeuAspLeuGlyGly..... 1594
      |||||
3118 TATACGCTTATCCGCAAGACGCGAG...TTCCGCTGCAATATCCGCT 3164
      |||||
1595 ...LysValIleSerAsnValIleGlyGlyThrLysAspThrAspAlaAl 1610
      |||||
3165 CAAGAACACAGAGCTTCCGAC.....AACTCGCAAGC 3199
      |||||
1610 asnValIleGlnIleuAsnGlnValArgAsnLeuLeuGlyLeuGlyAsn 1627
      |||||
3200 CGGGAACACAGAGCGCGCTTACGCGCAAAACGCAACTGCGCC 3249
      |||||
1627 laGlyAsnAspAsnAla.....AspGlyAsnGlnValAsnIleAlaAsp 1641
      |||||
3250 AACACACAGCGGAAAGAACACGCGCAAGCTTGAACGCGCTGATTGC 3299
      |||||
1642 IleLysLysAspProAsnSerLysSerSerAsnArgThrValIleLys 1658
      |||||
3300 GGGC.....GGCGCAATGCCACCGAAAGGCGAG 3328
      |||||
1658 saIaGlyThrValIleuGlyGlyLysGlyAsnAsnAspThrGlnLysLeu 1675
      |||||
3329 AAAGTGTGCGCAACCGCGCGGACGAGCGGGAATCCGCGCAT 3378
      |||||
1675 laThr.....GlyGlyIleGlnValIleGlyVal 1683
      |||||
3379 ATGAGCGGAGAGAGAAAAACGGGTGCAGGCGGATAAAGACCGC 3428
      |||||
1684 .....AspLysAsp..... 1686
      |||||
3429 CTTGCGCAACACCGCGAAGCGGAACCGCGCGCTACACCGCTTCC 3478
      |||||
1686 ..... 1686
      |||||
3479 CCCGCGCGCGCGCGCGCGGATTTCGCGAACCGGACCGCCACCG 3528
      |||||
1687 .....GlyAsnAlaAsnGlyAspLeuSerAsnValTrpValLysThr 1700
      |||||
3529 CAACCCCAA...CCGACGCGGACCTGATACGCCGTATGCCAATGCGG 3575
      |||||
1701 GlnLysAspGlySerLysLysAlaLeuLeuAlaThrLysAlaIleG 1717
      |||||
3576 TTGGTGAATTTTCGCGACGCTCAACAGCGTTTTCGCGTACAGAGC 3625
      |||||
1717 yGlnThrAsnThr...LeuThrAsnAsnProAlaIleAlaIle..... 1730
      |||||
3626 AATTGGACCGCGTGTTCGCGAA..... 3648
      |||||

```

```

1731 .....AspArgIleAsnGlnGlnGlyIleArgPhePheHisValAsnAsp 1745
      |||||
3649 .....GACCGCGCACCGCGCTTGGACAAACGG 3677
      |||||
1746 GlyAsnGlnIleProValValGlnGlyArgAsnGlyIleAspSerSerAl 1762
      |||||
3678 CATCCGGGACACCAACACATACGCTTCCGACAGATTTCCGCGCTACCGC 3727
      |||||
1762 a.....SerLysHisSerValAlaIleGlyPheGlnAlaLysAla 1777
      |||||
3728 AACAAACGACCTCGCGCAATCGGTATGACGAAAACTCGCGACGCGG 3777
      |||||
1777 spGlyGlnAlaIleValAlaIleGlyArgGlnThrGlnAlaGlyAsnGln 1793
      |||||
3778 CGCGTGCATCTGTTTTCGACACACCGACCGCA..... 3813
      |||||
1794 SerIleAlaIleGlyAspAsnAlaIleAlaThrGlyAspGlnSerIleAl 1810
      |||||
3813 ..... 3813
      |||||
1810 alIleGlyArgThrAsnValAlaIleGlyLysHisSerGlyAlaIleGly 1827
      |||||
3814 .....AACACTTCGACGACGCGCATCGCAC 3840
      |||||
1827 spProSerThrValLysAlaAspAsnSerThr.....SerValGlyAsn 1841
      |||||
3841 TCGCGACGCGCTTCCGCGCGCGCTTTCGCGCAATACGCGATCGCGAG 3890
      |||||
1842 AsnAsnGlnPheThrAspAlaThrGlnThrAspAlaIleGlyValGlyAs 1858
      |||||
3891 GTTCGACATCGGCATCAGCGCGCGCGGTATTAGTACGCGCACGCTTT 3940
      |||||
1858 n...AsnIleThrValThrGlnLysSerValAlaIleuGlySerAsn 1874
      |||||
3941 CAGACGCGCATCAGAGCAAAATCCCGCGCGCTGCTCATTCAGCAT 3990
      |||||
1874 erAlaIleSerAlaGly.....ThrHisAlaGlyThr 1884
      |||||
3991 CAGCGAATACCGCGCGCGCTTTCGCG..... 4017
      |||||
1885 GlnAlaLysLysSerAspGlyThrAlaGlyThrThrThrAlaGlyAl 1901
      |||||
4018 .....GGATTGCGCATGACACCGCATCGCGCGACCGC 4051
      |||||
1901 arThrGlyThrValLysGlyPheAlaGlyIleThrAlaValGlyAlaVal 1918
      |||||
4052 GCTATTTCGCAAAAGCGGATTCAGTACGAAACGTCATATC... 4098
      |||||
1918 erValGlyLysSerGlyAlaGlnArgArgIleGlnAsnValAlaIleGly 1934
      |||||
4098 ..... 4098
      |||||
1935 GluValSerAlaThrSerThrAspAlaValaAsnGlySerGlnLeuTyr 1951
      |||||
4099 .....GCCACCGCGCGCGCTTGCATTCACCGCT 4126
      |||||
1951 saIaThrGlnSerIleAlaAsnAlaThrAsnGlnLeuAspHisAlaGlyIle 1968
      |||||
4127 ACCGCGCGCGCATTAAGCAGATTAATTCATTAACCGCGGACACAT 4176
      |||||
1968 isGlnAsnGlnAsnLysAlaAsnAlaGlyIleSerSerAlaMetAlaMet 1984
      |||||
4177 TCCATTCACGCGCTTATTTGACCTGTCTATACGATGCGCTTCCGCGCA 4226
      |||||
1985 AlaserMetProGlnAlaIleProGlyArgSerMetValThrGlyIle 2001
      |||||
4227 AGTCCGAAGCGCGTCAATACCGCGGTATGCGGACGATTTCCGCA 4275
      |||||
2001 yIleAlaThrHisAsnGlyGlnGlyAlaValaIleValaIleGlyLeuSerLys 2018
      |||||
4276 ..ACCGCGAGTGCGAATG.....GGCTAAACGCGCAAAATC 4311
      |||||

```

```

2018 euSer1aspasn1yIntPval1Phe1ys1Ileasn1ySer1Ala1aspThr 2034
4312 AAGGTTTCACGCTGTCCCTCCAGCGTCCGCGCCGCAAGGG 4353
:::|||||
2035 Ingly1.....HisVal1Gly1Ala1Val1Gly 2043

seq_name: /cgn2_6/ptodata/1/lae/6B-COMB.pep:US-09-413-814-79

seq_documentation_block:
Sequence 79, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hottle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
EARLIER FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 79
LENGTH: 1213
TYPE: PR1
ORGANISM: Strangium celluloseum
US-09-413-814-79

alignment_scores:
Quality: 211.00 Length: 693
Ratio: 0.751 Gaps: 33
Percent Similarity: 40.548 Percent Identity: 24.820

alignment_block:
US-09-303-518d-653 x US-09-413-814-79 ..
Align seg 1/1 to: US-09-413-814-79 from: 1 to: 1213

2562 CAACGGCAATGTCCTCCAGCGATAGGCACTATTCATTGTAACA 2611
:::|||||
5 ArgArgArgAla1Ala1Pro1His1Leu1Arg1Gly1Ala1ArgPro1..... 17

2612 GCCGCTTACCGGAAATACGCGCGCAAGATACGCGCATTTACATTA 2661
|||:::|||||
18 .....Ala1ArgArgGly1Asp1Arg1Gly1Leu1Ala1Leu1ArgSer1 32

2662 AAGACAGCGAATGAGCGGTGCGCGGCGACGCAATTAAGCAATTTAAA 2711
|||:::|||||
32 sp1Arg1Ala1Arg1Ala1Arg1Arg1Gly1His1Gly1..... 42

2712 CTTTACACGCGCACCATTAACATCAATTCGCTATCGACACGATCGG 2761
43 .....Glu1Gly1Leu1Gly1Ala1Gly1 49

2762 CAGGGCGCAACCGGCAAGTGGCGAGATGCGCGCGCGCGCTTCCGCG 2811
|||||
49 YArg1Arg1Ala1Arg1Ala1Ala1Arg1Arg1Arg1Leu1Pro1Ala1Pro1Gly 66

2812 GTTCCCATTAATTCGTTACGCGCGCAACTTGGCAACAATTCGCTTCAA 2861
|||
66 roSerPro1..... 68

2862 CACGCTGACGATTAACGGCAATTTGAACGCTACAGCAACATTCGCTTTA 2911
69 .....Pro1Arg1Gly1Ala1Pro1Gln1Ala1Pr 76

2912 TGTGGAACTCTTCGCTACCGGACGCGCAATTTGAAGCTGGCGGAAGT 2961
|||||
76 oPro1Gly1Ala1Arg1Arg1Gly1Pro1Pro1Arg1Pro1His1Ala1Val1Gly1.... 91

2962 TCGGAAGGCACTTACACCTTGGCTGTCACATATCGCGCAACGACCGCT 3011
|||||
92 ..ArgArgAsp1Arg1Leu1Ala1Arg1Gly1His1Ala1Pro1Arg1Arg1 107

3012 AAGTTCGACGCAATTTGACGCTAGTGGAGGAAGAAGACACACACCGCTG 3061
:::|||||
108 Arg1Gly1Arg1Ala1Ala1Arg1Ala1Ala1Arg1Ala1..... 119

3062 CGGAATATCTTAATTTCAACCTTGCAAAAGCAACGCTGCGCGCGCA 3111
|||||
120 .....Pro1Ala1Ala1Arg1Arg1His1Gly1Arg1.... 128

3112 TGGCGTTATGACCTTATCCGCAAGACG.....CGAGTTCCG 3149
|||
129 .....Pro1Gly1Leu1Arg1Ser1His1Val1Arg1Val1His 139

3150 CTTGCATATTCGCTCAACAGACCTTTCCGCAAACTCGCGCAAGC 3199
|||
140 Gly1Ala1Ala1Gln1Gly1Arg1Asp1Arg1Pro1Arg1Arg1Gly1Gln1His1 156

3200 CGGAGCAACAGAGAGCGCGCTTGACGCGCAAAACGACACACTTGCCTGC 3249
|||
156 yPro1Arg1His1Gln1Pro1Pro1Leu1.....Arg1 165

3250 AACACACGCGCGCAAAAGCAACGCGCA.....AAGCT 3284
:::|||||
165 rArg1Arg1Pro1Gly1Gly1Pro1Gly1Ala1Arg1Ala1Val1Ala1Gly1Leu1Arg1Pro 181

3285 TGACGC.....GCTGATTCGCGCGCGCGCGCAATGCGACCGAAA 3322
|||||
182 Val1Gly1Leu1Arg1Arg1Val1Arg1Asp1Ala1Arg1Arg1Arg1Arg1Arg1As 198

3323 AGGCAAGAGTGTTCGCAACCGCGCGCGCGCGAGCGGCAAAATATTC 3372
|||
198 pPro1Arg1Pro1Gly1Pro1Arg1Leu1Gly1.....Ser1Gly1Ala1Leu1Ala1Arg1Ala 214

3373 GGCATTATGACGCGGAGAGCAAGAAAGCGGTGACGCGGATTAAGA 3422
|||
214 rGly1Ala1Arg1Ala1Gly1Asp1Arg1Val1Gly1Leu1Gly1Pro1Gly1Ala1Asp1Gly 230

3423 CACCGCTTGGCGAAGACAGCGGAGAGCGAAGCGCGCGCGCTACACCG 3472
|||||
231 Asp1Ala1His1Gly1Arg1Val1Pro1Arg1Arg1Gly1Arg1Pro1..Gly1Ala1Val1 246

3473 CCTTCCCGCGCGCGCGCGCG.....CGCGCGGATTTGCGCGCA 3513
|||||
246 lAla1Pro1Pro1Arg1His1Asp1Gly1Ala1Arg1Arg1Leu1Asp1Pro1Ala1Gln1Ala1Pro 263

3514 CGGACCGCGCAACCGCAACCGCAACCGCGAGCGGCACTGATCAGCGCTA 3563
:::|||||
263 rGSer1His1Pro1Arg1Gly1Leu1Pro1Arg1Ala1Pro1Arg1..... 274

3564 TGCATATGCGGTTTGAGTGAATTTTCCCGACCGCTCAACAGCGTTTTCG 3613
|||||
275 .....Arg1Gln1Pro1Arg1Arg1Arg1Asp1 282

3614 CGGTACAGAGCAATTTGACCGGCTGTTCGCGAAGACCGCGCAACGCG 3663
|||
282 G.....Gly1Val1Asp1Leu1Val1Asp1Arg1Pro1Pro1Asp1Arg1. 293

3664 GTTTCGACAGCGGCACTCGGAGACACCAACACTACCGTTGCGCAACATTT 3713
|||||
294 .....Gly1Arg1Arg1Pro1Gly1.....Val1Ala1Gln1His 302

3714 CGGCGCTACCGCGCAACAGCAACGACCTGCG..... 3743

```



```
303 ProLeuAArgProProAlaArgGluProAlaHisLeuArgAlaArgArgG1 319
|||||
3744 .....CCAAATCGGTATGCGAAGAACCTC 3768
|||||
319 yAlaGlyAlaValSerAspArgGlyProArgArg.....AspProH 333
|||||
3769 GGCAGCGGGCGCTCGGCATCCGTGTTTC..... 3797
|||||
333 ISArgArgAspArgArgAlaArgG1LeuAlaArgArgGlyAlaAsp 349
|||||
3798 .....GCACAAACCGACCGAACCCTTCAGCA.....CGGCA 3832
|||||
350 ProGlyAlaValProGlyAlaProHisAspArgArgAlaValGlnAs 366
|||||
3833 TCGGCACTCGGCAGCGCTGCCACGGTCCGCTTTCGGCAATACGG 3881
|||||
366 pArgArgProGlyProLeuArgArgArgAspHisArgAlaAlaArgA 383
|||||
3882 ..CATCG.....CAGGTTGCATCGGCATCAGCGC 3911
|||||
383 IAHIsArgProSerGlyGluAspProArgLeuProHisArgAlaArgArg 399
|||||
3912 GGGCGCGGGGTTTGTAGCGGCGACCTTTCAGACGCGCATCAGAGCAAA 3961
|||||
400 AspArgGlyArgProArgAlaAlaProLeuGlyArgAlaGlyArgG1 416
|||||
3962 TCCGCGCGCGCTGCTGCATTAGGCATTTCAGGCAAGATACGCGCAGT 4011
|||||
416 yGlyGluAspArgSerValGlyArgGluAla.....ProGlyArgV 430
|||||
4012 TTCGGGCGGTTTCGGCATCGACCGCACATCGCGCAGCGCTATTTCGT 4061
|||||
430 aArgArg.....ArgArgArgArgG1TPrArgArgAlaProArgLeuArg 445
|||||
4062 CCAGAA.....AGCGATTACCGATACGAAACGTCAATATCGCCACCC 4105
|||||
446 ProGluGluAlaAlaGlyValHisAspProArgGlyGlyArgArgProPr 462
|||||
4106 CGGGCGCTTCGATTCAACCGCTACCGCGCGCGCATTAAGCAGATTATTC 4155
|||||
462 oGlyAlaAlaAlaGluArgGluArgG1nGlyProArgArg..... 476
|||||
4156 TTCAAACCGCGCAACACATTCATCAGCCTTATTGAGCGCTGCTA 4205
|||||
477 ..ProAlaGlySerArgArgGlyArgAlaAla..... 486
|||||
4206 TACCGATGCCGCTTCGCGCAAGTCCGAACGCGCTCAATACCGCGGTAT 4255
|||||
487 ...ArgArgG1yArgArgProAlaAspArgAspArgAlaAlaHisArgLe 502
|||||
4256 TGGCGCAGGATTCGGCAAAACCCGCAATGCGGAATGGGGCGCTAAACGCC 4305
|||||
502 uGlyAla.....ArgG1yLeuAlaGlyArgGlyArgArgArgA 516
|||||
4306 GAATGCA.....AGTTTCACGCTGTCCCTCCAGCGTCCGC 4343
|||||
516 sPrArgG1nProLeuArgAlaArgLeuHisLeuAlaAlaProArgAlaArg 532
|||||
4344 .....CG 4345
|||||
533 AlAlaProAlaArgArgAlaHisArgGlyAlaArgAlaArgArgG1yAr 549
|||||
4346 CCAAGGGGCC..... 4355
|||||
549 gArgGlyAlaG1yArgValAlaAspArgProValProValProAspHisA 566
|||||
4356 ..GCAATTGAAGCGCAGCACAGCGCGG 4382
|||||
566 rGAlaTPrG1yAlaAlaAlaArgArgG1y 575
|||||
```

